UNIVERSITY OF HELSINKI
FACULTY OF AGRICULTURE AND FORESTRY

sRNA deep sequencing aided plant RNA virus detection in cultivated raspberries and molecular characterization of *Raspberry bushy dwarf virus* and *Black raspberry necrosis virus* from Finland

Fhilmar Raj Jayaraj Mallika

Master's Thesis

Master's in Plant Production Sciences (Plant Virology)

Department of Agricultural Sciences

University of Helsinki

May 2018

Tiedekunta/Osasto - Fakultet/Sektion – Faculty	Laitos - Institution – Department
Faculty of Agriculture and Forestry	Department of Agricultural Sciences

Tekijä - Författare – Author
Fhilmar Raj Jayaraj Mallika

Työn nimi - Arbetets titel – Title
sRNA deep sequencing aided plant RNA virus detection in cultivated raspberries and molecular characterization of *Raspberry bushy dwarf virus and Black raspberry necrosis virus* from Finland.

Oppiaine - Läroämne – Subject
Plant Virology

Työn laji - Arbetets art – Level	Aika - Datum – Month and year	Sivumäärä - Sidoantal – Number of pages
Master's Thesis	May 2018	61

Tiivistelmä - Referat – Abstract
Plant RNA silencing machinery cleaves exogenous nucleic acids into 21, 22 and 24nt siRNAs, which constitute the core of basal defence against RNA/DNA virus infection. sRNA deep sequencing and assembly of 21-24nt size reads from virus-derived siRNAs (vsiRNAs) details plausible information of viral genomes targeted, following sequence homology search in virus databases. Pooled RNA sample marked (HXR-2) from 21 raspberry cultivars subjected to sRNA deep sequencing and vsiRNA read alignment revealed the presence of *Raspberry bushy dwarf virus* (RBDV) and *Black raspberry necrosis virus* (BRNV). Countercheck using VirusDetect, a sensitive virus finding software also confirmed these raspberry infecting plant RNA viruses through contig assembly. Read size calculation attested 21nt and 22nt vsiRNA read length as leads in alignment to reference that covered genomic regions in both viruses. Viral genome mapping displayed vsiRNA distribution regions, RBDV genomes distinctly masked from 5' end, towards coding regions in RNA-1 and RNA-2, scarcely towards 3' end. The putative stem-loop occurring 3'UTRs were figured as least covered regions. BRNV genomes mapped fractionally at 5'UTRs, and densely at 3'UTRs including a vsiRNA 22nt size hotspot position. RdRp encoding region in RNA-1 and small coat protein (CPs) encoding region in RNA-2 were markedly not spanned by vsiRNAs. RT-PCR for individual samples validated RBDV presence in 5 cultivars and BRNV in 6 cultivars. Molecular characterization following Sanger sequencing distinguished virus isolate diversity, RBDV sequence phylogeny showed limited variability in the CP encoding region among geographical isolates. BRNV CPs encoding region showed 25% nucleotide non-identity among cultivar isolates, tree phylogeny supported the interchange of isolates among cultivars and wild raspberries in Finland. The study highlights novel virus detection tool and significant vsiRNA benchmark profile for discovering RNA viruses. Remarked RNA recombination in BRNV CPs encoding sequence is moreover surmised.

Avainsanat - Nyckelord – Keywords
Black raspberry necrosis virus, BRNV, RBDV, *Rubus idaeus*, *Idaeovirus*, vsiRNA, 21nt, 3'UTR, RNA silencing, RNA virus

Säilytyspaikka - Förvaringsställe – Where deposited
Department of Agricultural Sciences and Viikki Science Library

Muita tietoja - Övriga uppgifter – Further information
Supervised by: Prof. Jari P.T. Valkonen

CONTENT TABLE

1 INTRODUCTION

Plant quarantine agencies undertake the duty of averting exotic viruses and viroids at border level for risk minimization in pathogen entry, also considering sustainable cultivar production and native plant species (Barrero et al. 2017). As plant materials move, introduction risk of new and evolving viral pathogen increases, plant biosecurity is the need for securing cultivars from such agents causing disease epidemics. This is expanded to scrutinize contaminated plant products moving at national and international level (Martin et al. 2016). Virus indexing in plant germplasm is tedious task at international level, for it takes six months in case of Sweet potato (*Ipomoea batatas* L.) to complete essential plant virus probing methods (Kreuze et al. 2009). Nonetheless, small RNA (sRNA) omics based detection of plant viruses and viroids is being effectuated to detect and identify evolving as well as new agents of viral diseases (Seguin et al. 2014). The method relies on the enrichment of anti-viral silencing product, virus-induced small-interfering RNA (vsiRNA), by plant cellular defence pathway that naturally encounters viral nucleic acid. The assessment and utilization of this technology is underway for universal diagnosis of viral pathogens (Kreuze et al. 2009, Roy et al. 2013, Valkonen 2014).

Red raspberry (*Rubus idaeus* L.) and Black raspberry (*Rubus occidentalis* L.) belong to the genus *Rubus*, widely grown for taste, nutrients and its presence of antioxidants (Tzanetakis 2012, Martin et al. 2016). Vegetative propagation is the method employed to raise new scions for cultivation purposes. Plants are perennial in nature, living up to 15 more years from first planting date, despite challenges from biotic and abiotic factors. Varieties developed are prone to plant viruses during all growth stages, and disease epidemics are accelerated by vectors or natural factors. Clonal propagation method makes the nuclear stock to transmit viruses to newly developing primo canes, leading to infected plantings over a large area (Martin et al. 2013). In *Rubus* species, asymptomatic development of diseases by viruses and virus complex forms the barrier to characterize the agent by purification. The less concentration of viruses in tissues following single infection also conform the challenge to detect and characterize new viruses or strains of known viruses. Coordinated approach of pathogen control is needed for effective control of viruses, and the best method is to start with certified virus free resistant cultivars (Tzanetakis 2012). For the demand for high quality berries, *Rubus* plants are being traded, the way by which viruses are being transmitted

across the world. In the United States, *Rubus* virome developed due to mislabelling of an indicator plant for virus testing. Black raspberry cultivar 'Munger' was sold for many years as indicator plant, and showed no viral symptoms for many recently detected viruses. However, two synergistic viruses lately identified had been causing severe epidemics, leading to the decline of Blackberry plants, when different viruses were tested with 'Munger' as indicator plant. Consequently, screening for virus symptoms with one cultivar for many known and new viruses through Biological Indexing (BI) advantaged the emergence of Blackberry yellow vein disease complex in southern states. The other example is the trans-trade of *Rubus sp* that spread the viral pathogen, *Raspberry bushy dwarf virus* (RBDV). This pollen-borne virus got distributed worldwide through propagated canes, and too its genetic diversity limit was correlated. It was also reported to infect grapevine, further studies on host range were not described (Valasevich et al. 2011).

In Finland, RBDV was found in 42 clones of red raspberry kept under field-maintenance, knowing this meristem invading virus persistence, working group of berries and fruits had to develop novel therapeutic methods for elimination, equally to commercialize breeding line. It is also prevalent in wild plants of *Rubus*, reported presence in Arctic bramble (*Rubus arcticus* L.), a plant growing well in north temperate zones (Kokko et al. 1996, Wang et al. 2008, Lemmetty et al. 2011). sRNA deep sequencing has been utilized in Finland to detect novel plant viruses in raspberries. This technology had provided the first report of *Raspberry leaf blotch virus* occurrence in a susceptible cultivar in Finland, and later virus distribution in wild raspberries (Bi et al. 2012, Parikka et al. 2016). The short read assembler Velvet is selected as the tool to assemble contigs preciously following read alignment parameters (Bi et al. 2012, Nordenstedt et al. 2017).

Aphid borne *Black raspberry necrosis virus* (BRNV) concentration is usually found low in raspberry cultivars that indicator plants cannot distinguish strain variability, RT-PCR testing of this virus at different geographical regions is under suspicion. Nevertheless, methods to detect all available strain of this virus is an essential task (McGavin et al. 2010, Martin et al. 2013). To check virus persistence in certified *Rubus* germplasm in Finland, RNA sample pool (HXR-2) from 21 raspberry cultivars had been subjected to sRNA deep sequencing. Previously, Velvet assembly developed contigs provided detection results for viruses that also included two (+) RNA viruses:

RBDV and BRNV. The initial objective of this study was set to detect these viruses by RT-PCR in individual raspberry cultivars of sample pool for sequence analysis. The molecular diversity of BRNV, a low titre-virus in raspberry was anticipated from previous studies (Halgren et al. 2007, McGavin et al. 2010, Martin et al. 2013) and isolate accessions from Finland (unpublished). While the work was in procession to characterize BRNV and RBDV isolates, a novel virus finding software was released, which can rapidly produce results of viruses, and *de novo* assemble genomes of novel viruses. Chipster, user-friendly bioinformatic toolkit under CSC Finland (Kallio et al. 2011) recently included VirusDetect (Zheng et al. 2017) in its list of programmes. Meanwhile, (Barrero et al. 2017) showed sorting read size of sRNA profiles that produced straight-forward results for virus detection. Out of 21 multiplexed samples obtained from quarantine agencies in Australia and New Zealand, 13 showed positive for 18 viruses and 3 viroids including RBDV. The workflow simplified the task of detection using bioinformatic pipeline, and RBDV genome was covered 97.89% by 4 contigs, when subset containing vsiRNA size 21-22nt was used to assemble *denovo* genomes of viruses. VirusDetect evaluation for release had demonstrated (Zheng et al. 2017) the same accuracy of identification using sRNA profile. In the light of these details and easy accessibility to VirusDetect, the objective of this study was reformulated to reconfirm the presence of RNA viruses in raspberry samples of HXR-2. Furthermore, genome coverage analysis of vsiRNA profile (read size) to map regions in RBDV and BRNV genomes has been added to this study.

2 REVIEW OF LITERATURE

2.1 Plant RNA virus replication

Plant viral genomes are grouped according to nucleic acid constituent, structure and polarity, and so far single stranded (ss) (+) RNA virus, ss (-) RNA virus, dsRNA virus, ssDNA virus and pararetro (dsDNA-RT) virus are being listed as invading alien nucleic acids of plant cells (Carbonell et al. 2016). The preponderant group comprises of non-enveloped yet encapsidated positive-sense ssRNA genome which replicates in cell cytoplasm. Coat protein (CP) molecules indemnify all available genomes tightly, and are multifunctional in nature. Systematically coordinated transcription, replication, translation and protein disassembly occur for successful infection in host cells (Ivanov & Mäkinen 2012). Viral Replication Complexes (VRC) are re-assembled organelle like structures by viral RNA, viral membrane-bound/auxiliary proteins along with host membrane associated proteins, where inside replication event takes place. (Carbonell et al. 2016). Generally, infectious (+) genomic RNA (gRNA) function according to its size, number and structure of RNA segment/segments (Ivanov & Mäkinen 2012, Carbonell et al. 2016). The monopartite RNA genome consists of one gRNA, *Potyviridae* and monopartite *Secoviridae* family viruses with such genomes operate through a strategy to express individual yet multifunctional proteins prior to replication. This strategy is described as polyprotein processing, when coding regions are initially translated as single polyprotein, further cleaved by proteases of viral origin (Ivanov & Mäkinen 2012, Thompson et al. 2017). Cleaved proteins facilitate replication after positioning towards cellular membranes like Endoplasmic Reticulum (ER) to VRC connections (Carbonell et al. 2016).

The replication process is executed by viral enzyme complex known as RNA dependent RNA polymerase (RdRp). This error-prone polymerase is first translated from part of 5’ Untranslated region (UTR) and it synthesizes negative (-) strand from the gRNA for progeny RNA production (Carbonell et al. 2016). Multipartite viruses harbour two or more gRNAs, enhance replication and transcription by *cis-acting* or *trans-acting* on regulatory elements present, at points also contain potential signal for (–) strand or sub-genomic (sg) RNA synthesis (Ivanov & Mäkinen 2012). The widely investigated monopartite and multipartite viruses are *Tobacco mosaic virus* (TMV), *Tomato bushy stunt virus* (TBSV), *Brome mosaic virus* (BMV) and *Alfalfa mosaic*

virus (AMV), in these viruses replication events are around the same corner (Wu et al. 2013, Chujo et al. 2015, Carbonell et al. 2016). In BMV, replication and CP assembly would occur simultaneously. Inside VRCs, + ssRNA viruses replicate by forming replication intermediates of perfectly double stranded, also transcribes (mRNA synthesis) by base paring with regulatory regions (*cis*-elements) present in UTR regions (Carbonell et al. 2016). The RNA molecule folding develops secondary structures like stem loops (SL)/hairpin through RNA: RNA interactions. The non-coding 5'UTR, regulatory elements and 3'UTR serve as axes for viral RNA replication process (Ivanov & Mäkinen 2012, Wu et al. 2013, Chujo et al. 2015, Carbonell et al. 2016).

Re-programing of virus occurs when CP disassembled viral RNA co-translationally replicates when RdRp leaning on the 5'UTR of genome. In TMV, after negative (-) RNA synthesis following replication, sgRNA molecules are expressed from negative strand templates (Chujo et al. 2015). (-) RNA synthesis initiates from sgRNA promotors located at 3' terminal and proximal sequences in BMV and TBSV (Wu et al. 2013, Carbonell et al. 2016). Genomic features like tRNA-like structure (TLS) and 3'proximal pseudoknot affect complementary (-) RNA synthesis in *Tobamovirus. Tobamovirus* replication proteins bind to 3'UTR region that contains TLS to synthesis complimentary strand. Deletion of 3'UTR regions can impair replication and infectivity in all above mentioned viruses (Chujo et al. 2015). SL of BMV at 5' end of RNA-1 and RNA-2, also found in RNA-3 intergenic recombining region known as B-box, which interacts with replication associated proteins recruiting replication (Carbonell et al. 2016). In TMV replication proteins bind to 5'end for strong template selection and to switch between translation and replication, especially when ribosome and RdRp scanning the same template, yet from opposite directions (Chujo et al. 2015).

2.2. Plant RNA virus recombination

In nature, a virus evolves more rapidly than its counter host resulting diverse RNA population replicates, and when such population reach an equilibrium, it can be termed as 'Quasispecies' (Schneider & Roossinck 2001, Martin et al. 2016). Normally, recurring in a single host would keep the quasispecies cloud of RNA virus. *Cucumber mosaic virus* (CMV) host range is more than 1000 species and its quasi cloud remains

not constant, and this is a factor of virus-host interactions that permissive host might allow high level of viral population diversity, or a non-permissive host would decelerate viral replication, and diversity changes due to defense counteraction. The diversity is stimulated by two factors, (I) polymerase error and (II) recombination (Schneider & Roossinck 2001, Barr & Fearns 2010, Quito-Avila et al. 2012). In the event of RdRp /Polymerase error during elongation, the internal or coding regions of viral RNA encounter damage, and the virus repairs by recombination. It was reported that each molecule of BMV RNA replication resulted progenies during one recombination event (Barr & Fearns 2010). Moreover, internal damage can also occur as a result of alkylating agents, and for repairing coding regions few viruses carry domains related to AlkB protein, which can repair DNA and RNA methylation damage (Van den Born et al. 2008, Barr & Fearns 2010). In the event of RdRp/Polymerase error during replication initiation or termination, nascent or abortive transcripts are generated by RdRp for viral recombination repairing terminal damage (Barr & Fearns 2010, Quito-Avila et al. 2012). For example, in *Turnip crinkle virus* satellite RNA C (satC), loss of 3'terminal regions were repaired by its helper virus *Turnip crinkle virus* (TCV). Primers generated from TCV template was reported as evidence for this homologous recombination. Non-homologous recombination by non-template primers, generated *denovo* by TCV RdRp in a single step also repaired 3'end of satC. This was too shown for CMV Sat by its helper virus CMV at 3'end (Barr & Fearns 2010).

2.3 Anti-viral RNA silencing

Dispensable RNA molecules in eukaryotic cell cytoplasm are silenced by the action of cellular endoribonuclease belonging to conserved RNAase-III family (Weinheimer et al. 2015, Ghoshal & Sanfacon 2015). Consequently, RNA silencing in plants outlines basal defense from endogenous and exogenous dsRNA or structured aberrant RNA molecules (Kreuze et al. 2009, Donaire et al. 2009, Pumplin & Voinnet 2013). During replication process double stranded (ds) intermediate pattern occurs; concurrently during recombination events, when RdRp switching templates. The loss of vulnerable terminal regions effect viral RNA repair, and are also related to cleavage by cellular nuclease activity (Barr & Fearns 2010). RNA viruses induce silencing during replication and are targets at post-transcriptional level, this is collectively known as Post Transcriptional Gene Silencing (PTGS). The signature molecule of PGTS

antiviral-silencing is a non-coding (nc) short-interfering RNA (siRNA) of 21 and 22nt in size/length. These are follow-up products of DICER-Like (DCLs) ribonucleases/proteins that consist of RNAse III domain. DCL-4 dices for 21nt size and DCL-2 for 22nt siRNA, with a 2' overhang at 3'OH of the duplex (Pumplin & Voinnet 2013, Weinheimer et al. 2015, Liu et. al 2016) (Figure 1). Complementing viral intercession of siRNA generation, the term virus-derived siRNA (vsiRNA) is favourably denoted, which also target homologous viral mRNAs (Pumplin & Voinnet 2013).

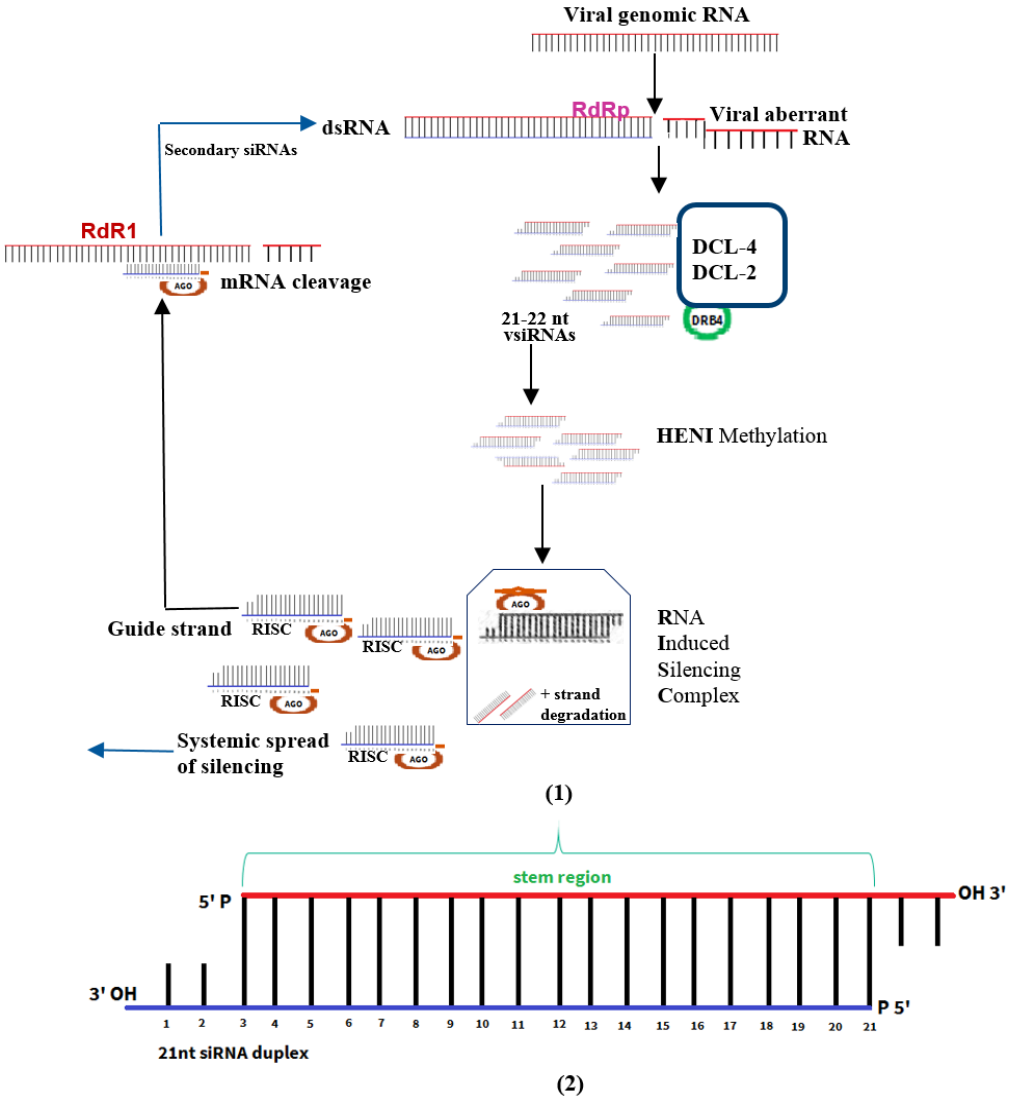


Figure 1. PTGS anti-viral immunity in plants. **(1)** Production of vsiRNAs following plant RNA virus interaction. **(2)** General structure of siRNA duplex.

DCLs are assisted by cofactors known as DRB (Double stranded RNA binding Protein). For example, DCL-4 associates with DRB-4 and together action dices long dsRNA and hairpin loop structures into 21nt vsiRNAs (Ghoshal & Sanfacon 2015). A

structure association is attributed for recognizing the target by DCL proteins (Donaire et al. 2009, Wu et al. 2013). The cleaved 21nt and 22nt vsiRNA terminal nucleotide is 2'-O- methylated for stability through HEN1 (Huna Enhancer 1) and are loaded by DCL/DRB into a cleavage complex, denoted as RISC (RNA induced silencing complex). Inside action of RISC is perfected by Argonaute protein (AGO) that contain PIWI domain of RNase H type, slicing the sense (+) vsiRNA duplex derived from RNA viruses is shown as AGO2 and AGO1 mediated action. The (-) negative strand of the sliced vsiRNA molecule, from now known as 'guide strand' accessible to AGOs, is directed for sequence homology search towards complementary sequences that are either viral mRNAs or replicating intermediates, for sequence-specific cleavage (Pumplin & Voinnet 2013, Weinheimer et al. 2015).

Unlike vertebrates and insects, plants are able to enrich vsiRNAs of viral infection to higher thresholds for efficient systemic resistance, and biogenesis of 21nt, and 22nt vsiRNAs were found in abundance following RNA viral infection, to resistance signalling throughout tissues (Donaire et al. 2009, Pumplin & Voinnet 2013). This is facilitated by cellular RNA dependent RNA polymerase (RDR1/RDR6) that initially recognizes the cleaved mRNA targeted by RISC associated and vsiRNA guided AGO, and amplifies the molecule into replicating dsRNA form. SGS3 (SUPPERSSOR OF GENE SILENCING 3) is the co-factor of RDR6 for amplification of short or long cleaved viral mRNA transcripts. The priming less amplification of RDR1/6 consequence RNA replication, is recognized by DCLs to produce secondary vsiRNAs. The plant vascular tissue (phloem) system channels vsiRNAs towards tissues ahead of viral RNA movement, for naïve tissues can autonomously counteract RNA products of cognate virus (Ghoshal & Sanfacon 2015, Pumplin & Voinnet 2013). The general events of PTGS silencing pathway is shown (Figure 1).

The DNA and dsDNA-RT replication in nucleus are also targeted by plant endonucleases, however the mode of action is diverse. It occurs at places where host gene expressions, Transposable elements (TEs), and inverted gene repeats are tightly controlled at Transcriptional level (Pumplin & Voinnet 2013). Therefore DNA, dsDNA-RT virus genomes and mini chromosomes undergo RNA dependent DNA methylation (RdDM), also since those produced viral transcripts are outcomes of transcription by cellular RNA polymerases like Pol IV and Pol V. Here, DCL-3 dices the transcripts of DNA viruses amplified by RDR2 into 24nt vsiRNAs, subsequent

loading into AGO4. Methyltransferase, MET1, and other methyl group enzymes methylate and maintains methylation of homologous DNA and arrests genes by blocking the accessibility for Polymerase II, also virus integrated sequences into host genome can also undergo Transcriptional silencing. In addition to that 21nt and 22nt vsiRNAs are also generated from RNA transcripts of viruses with DNA genomes, possibly interfering with PTGS in cytoplasm (Pumblin and Voinnet 2013, Ghoshal & Sanfacon 2015). Non-coding, autonomous replicating viroids use cellular pol II for replication and are targeted by all DCLs generating 21, 22, and 24nts vsiRNA duplexes, but viroids are poor target for RISC harbouring AGOs due to its compact degree structure (Seguin et al. 2014, Hohn 2015).

As principle, this sequence-specific slicing of viral mRNAs would protect the plant from any secondary infection of cognate or a related virus with considerable sequence similarity. The plant recovering by antiviral silencing show recovery phenotype, which is also shown following infection with RNA and DNA viruses. This is facilitated by systemic spread of silencing signal and reduced virus threshold in tissues. However, symptom induction and recovery are outcomes of weak or strong plant virus-interactions (Ghoshal & Sanfacon 2015). In an unrelated way, due to sequence homology, vsiRNAs can also enter host-gene silencing pathway that could induce radiant symptoms in plants by silencing host genes. CHLI (Chelatase subunit) gene is responsible for chlorophyll biosynthesis in plants, following *Cucumber Mosaic virus* Y-satellite RNA (Y-Sat) infection in *Nicotiana tabacum* L., one single molecule antiviral silencing product of vsiRNA derived from Y-Sat targets CHLI mRNA for cleavage, thus shows severe yellowing symptoms in leaves. In *N. tabacum* relatives *Nicotiana clevelandi* and *Nicotiana debneyi* the CHLI gene homology is different from Y-Sat that it cannot induce yellowing symptoms (Pumplin & Voinnet 2013, Ghoshal & Sanfacon 2015).

2.4 Viral suppression of RNA silencing

Viruses counter-act RNA silencing by expressing proteins known as Viral Suppressors (VSR) (Pumplin & Voinnet 2013, Weinheimer et al. 2015). The widely checked multifunctional Helper-component protease (HCPro) protein of *Potyviridae* and VSR protein p19 of TBSV impounds vsiRNA of 21nt and 22nt size. This prevents vsiRNA loading into RISC complex, allows viral movement and symptom induction (Ghoshal

& Sanfacon 2015). The non-membrane bound 130KDa protein subunit of TMV with helicase domain known as P126, hinders HEN1 methylation of vsiRNA duplexes thus making unstable for RISC incorporation (Pumplin & Voinnet 2013). The DRB4/DCL4 21nt vsiRNA production activity is counteracted by direct interference of TAV (transactivatior or P6) protein of *Cauliflower mosaic virus* (CaMV) (Pumplin & Voinnet 2013). The Argonaute (AGOs 1 & 2) degrading VSRs are *Potato virus X* (PVX) P25 protein, CMV multifunctional 2b protein, *Potato leafroll virus* (PLRV) P0 protein and *Tomato ringspot virus* (ToRSV) CP protein (Ghoshal & Sanfacon 2015). Both 21nt and 22nt, also 24nt vsiRNA duplexes are reduced to inactive 14nt size by strong endoribonuclease activity of viral RNaseIII expressed from the genome of *Sweet potato chlorotic stunt virus* (SPCSV), which promotes viral synergism with *Sweet potato feathery mottle virus* (SPFMV) (Pumplin & Voinnet 2013, Weinheimer et al. 2015, Ghoshal & Sanfacon 2015). VSRs also contain potential signaling/motifs that can target cell membrane organelles to circumvent the line of RNA silencing. Remarkable information comes from VSR of *Peanut clump virus* (PCV), P15 which could bind more 22nt than 21nt vsiRNAs to peroxisomes by its PTS1 (peroxisomal targeting signal), impairing the systemic spread of silencing signal. (Daros 2017, Incarbone et al. 2017).

Proteinaceous VSRs are selectively being studied from considerable viruses of long host range, precursory investigation reveals the existence of RNA mediated suppression (Hohn 2015). This evidence proceeds from long non-coding (lnc) RNA of both RNA and dsDNA-RT virus, and show VSR dependent and independent mode of action in silencing suppression (Hohn 2015, Flobinus et al. 2016). The virus responsible for ‘Rhizomania’ through *Polymyxa betae* (protozoa) in sugar beet, contains three gRNA, and the expression of VSR protein p14 from RNA-2 is needed to move long-distance by counteracting RNA silencing RDR6 pathway. RNA-3 produces a long ncRNA3 by 5’-3’ exoribonuclease activity of XRN1. The core 20nt sequence of ncRNA3 is also required for long-distance movement of virus and this RNA vital sequence motif is known as ‘coremin’. In the absence of VSR p14 protein, ncRNA3 interferes antiviral silencing pathway, allowing systemic spread of virus. In *Red clover necrotic mosaic virus* (RCNMV) too this type of ncRNA production is shown (Flobinus et al. 2016). The showpiece action of non-coding 600nt long leader sequence, known as 8S of CaMV 35S RNA competes for engaged AGOs, while

forming a dsRNA along with anti-8S for that origin is unknown (Hohn 2015). This highly structured region of pre-genome separately produces 82% vsiRNA of 21, 22 and 24nts involving all DCLs in contrast to other parts of genome (Hohn 2015, Ghoshal & Sanfacon 2015). A sensitive method of vsiRNA analysis, known as 'sRNA deep sequencing' clearly showed 8S action along CaMV 35S RNA genome, further study also unclosed 8S decoy type vsiRNA production in another dsDNA-RT virus (Hohn 2015).

Size groups of 21, 22nt and 24 vsiRNA are generated from cytoplasm or nucleus following viral replication or suppression of silencing (Ghoshal & Sanfacon, 2015). Plant viral genome mapping would provide fractional details of VSR activity or viral elements showing effect, which are also virus specific to strain specific of silencing. (Donaire et al. 2009, Ghoshal & Sanfacon, 2015, Hohn 2015). Production of vsiRNA is reported from all regions of plant viral genomes, adding from sense and anti-sense strands (Donaire et al. 2009, Seguin et al. 2014).

2.5 Plant Viruses of *Rubus*

Rubus serves as an infection harbourage for 40 known plant viruses, more than 15 viruses particularly affecting raspberry and black raspberry had been listed (Martin et al. 2013). Raspberry mosaic disease (RMD) is a virus complex/ viral synergism, which includes *Raspberry leaf mottle virus* (RLMV), *Rubus yellow net virus* (RYNV) and *Black raspberry necrosis virus* (BRNV) (Martin et al. 2013). These viruses are transmitted by the aphid *Amphorophora agathonica* to healthy plants in North America, and large raspberry aphid *Amphorophora ideai* transmits viral infection in Europe (McMenemy et al. 2012, Martin et al. 2013). The other important disease is Raspberry crumbly fruit caused by *Raspberry bushy dwarf virus* (RBDV) and its virus complex (Martin et al. 2013).

2.5.1 *Raspberry bushy dwarf virus*

RBDV is a major viral pathogen causing severe yield reduction and reduced plant growth, and is found to occur wherever *Rubus* plants are found, especially affecting red raspberry production (Valasevich et al. 2011, Martin et al. 2013, Isogai et al. 2015, Martin et al. 2016). It is placed under the genus *Idaeovirus* (Family *unassigned*) due to its distinguishable characteristics from all other plant RNA viruses. The virus has

narrow host range on cultivated and wild *Rubus*, found infecting raspberry and blackberry. It is usually symptomless in most cultivars of raspberry, but exhibit symptoms such as radiant yellow pattern for ‘yellows disease’, and severe dwarf for associated disease decline known as ‘crumbly fruit’, in sensitive cultivars. The crumbly fruit disease induced would cause severe yield loss by means of reduced drupelets. Symptoms include leaf yellowing, reduced cane height and fruit reduction, both qualitatively and quantitatively. Later studies had uncovered that a viral complex cause this disease. Due to severe economic loss, infected plantings were recommended for removal to install virus free plants. (Martin et al. 2013, Quito-Avila et al. 2014b). The virus is transmitted by pollen and seed, both horizontal and vertical transmission of virus is feasible for spreading the disease (Isogai et al. 2015). The pollen is systemically infected first and successful vertical transmission of RBDV occurs, when infected pollen tubes containing virus particles are introduced into embryo sacs, while the pollen stays viable but the drupelet formed is aborted. The concentration of the virus is high in infected pollen tubes, and are further transmitted by wind and other pollinating insects. (Isogai et al. 2015).

Figure 2. Bipartite genome organization of RBDV RNA-1 and RNA-2. Both coding and non-coding genomic regions are shown in boxes. RNA-1 encodes Replicase proteins and putative protein p12 coding region in RNA-1 is marked with question mark. RNA-2 encodes Movement protein, Capsid protein is expressed from separate sub-genomic RNA, upstream of RNA-2 3’ Untranslated region.

Virions are quasi- isometric particles, 33nm in diameter and contains 24% RNA. The bipartite RNA genome molecules are bicistronic in expression. RNA-1 is 5449 nts long and RNA-2 length is 2231 nts. A sub-genomic, encapsidated RNA-3 is present as derivative from RNA-2. This 946nt long molecule is monocistronic, expressing the viral coat protein (CP) from 3’end (Harrison & Murant 1996). RNA-1 consists of two over lapping Open Reading Frames (ORF), ORF-1a and ORF-1b having sequence similarity. The ORF-1a encodes a polypeptide of 190kDa, which has sequence similarities or conserved motifs of Methyl transferase (MTR), Helicase (Hel) and viral RNA dependent RNA polymerase (RdRp). Putative p12 protein was positioned in RNA-1, but not verified experimentally. The RNA-2 consists of two ORFs, ORF-2a and ORF-2b, from which the derivate of CP occurs as a sub-genomic RNA. The ORF-

2a encodes 39kDa putative Movement protein (MP) which has sequence similarities to other RNA viruses. ORF-2b encodes the CP of 30kDa, verified experimentally (Harrison & Murant 1996, McFarlane et al. 2009). RBDV genome organization is shown (Figure 2). The CP of this virus was reported as the key element in inducing infection in host plants. (MacFarlane et al. 2009). In one RBDV isolate, CP encoding partial sequence was found being inverted or copied to the 5' terminal of the RNA-2. (Quito-Avila et al. 2014b). Two important viral strains were reported, S isolate (Scottish) and RB (Resistance-breaking) strain that found in both North America and Europe. Genetic resistance against RBDV is confined by a single resistance *Bu* gene, but RB strains are found breaking this natural source of resistance (McFarlane et al. 2009, Valasevich et al. 2011). RBDV is commonly found with an aphid transmitted virus, BRNV that too is also an agent in 'bushy dwarf symptom', hence occasionally causing raspberry shoot proliferation. (Martin et al. 2013). The virus also synergies with other *Rubus* infecting viruses like RLMV in virus complex, where RBDV increased titres were reported. (Tzanetakis 2012, Quito-Avila & Martin 2014a). In wild, RBDV is found infecting *Rubus idaeus* var. *sirigosus*, *R. multibracteatus* H. Lev. & Vaniot, *R. oocidentalis* L., *R. parviflorus* Nutt., and *R. molaccanus* L. (Martin et al. 2013). In Finland, RBDV presence was confirmed in 1993, infecting red raspberry genotype and soon found in raspberry cultivars and wild raspberries. It is known locally as Vadelman kääpiökasvuvirus, the first infection in arctic bramble had been reported from Finland in 1996 (Kokko, et al. 1996, Lemmetty et al. 2011, Parikka et al. 2016). The suggested way to control this virus is to plant virus free plants or to use disease resistance cultivars, and considering protected cultivation techniques. The natural infection to seed remains a great hindrance for RBDV control measures. The development of transgenic resistance in raspberry cultivar 'Meeker' had been reported (Mathews et al. 2004). Virus free plants can be generated by 'cryotherapy', RBDV was shown as the example for virus elimination from meristematic cells (Wang et al. 2008).

2.5.2 *Black raspberry necrosis virus*

Black raspberry necrosis virus (BRNV) is an aphid-borne + ssRNA virus found infecting plants of *Rubus* sp, both wild and cultivars (Martin et al. 2013). The pathogen is recalcitrant to characterize, and increased titres were observed only at times

inoculated with viruses of *Sobemovirus* genus. Probing for BRNV using Electron Microscopy (EM) and RT-PCR quite revealed a new *Sobemovirus*, including complete genome sequence (McGavin & McFarlane 2009).

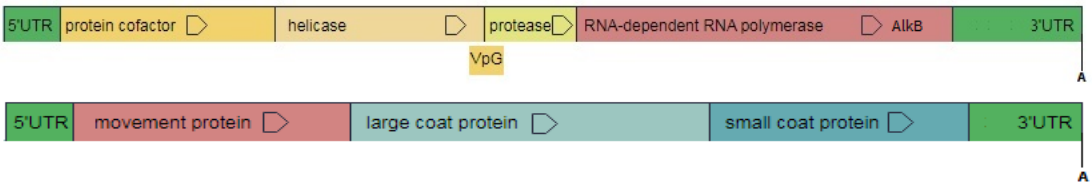


Figure 3. Bipartite genome organization of BRNV RNA-1 and RNA-2. Poly A-tail is marked as A at the 3'end. Both coding and non-coding genomic regions are shown in boxes. RNA-1 encodes Replicase and polyprotein processing protein. Putative VPg is shown in genomes. RNA-2 comprises of movement protein encoding and two coat protein encoding regions.

BRNV is an important component of RMD complex and is transmitted by *Amphorophora sp*, severe mosaic symptom like appearance to death of plants in cultivated black raspberry was reported. This virus is found symptomless or causes latent infection, but induces severe symptoms in cultivated black raspberries, unlike red raspberry cultivars. It is the causal agent behind Black raspberry decline (BRD) in the United States (Halgren et al. 2007). Spherical shaped virus particles cover its bipartite genome, two RNA molecules of 7572nt and 6350nt excluding the poly-A tail (McGavin et al. 2010). The infectious single RNA may not initiate infection systematically, and both RNA molecules are vital for viral replication and systemic movement (Thompson et al. 2017). Translated polyprotein is large in size undergoes 3C-like protease cleavage, RNA-1 Replicase protein consists of nucleoside triphosphate (NTP)-binding proteins (Helicase), 3C-like proteinases (Pro) and RdRp. The conserved domains in RdRp codons represents a remarkable DNA and RNA repairing AlkB protein of oxygenase superfamily (Halgren et al. 2007, Thompson et al. 2017). Putative VPg (Viral genome linked) protein is attached covalently to the 5' end of genomic RNAs. RNA-2 encodes the movement protein (MP) and two forms of structural coat proteins, large coat protein (CP1) and small coat protein (CPs). Identical nucleotides at 3' UTR of RNA-1 and RNA-2 is reported. Both genomes are poly-adenylated, at the 3'termini (McGavin et al. 2010, Thompson et al. 2017). BRNV genome organization is shown (Figure 3). The protein encoded by isolates from US and UK were also found to be diverse at nucleotide level (McGavin et al. 2010). BRNV is ungrouped under the family *Secoviridae* and phylogenetically aligns with another unassigned, *Strawberry mottle virus* (SMoV) and *Satsuma dwarf virus* (SDV), type

member of genus *Sadwavirus* (Halgren et al. 2007). ER- membrane associated VRC based replication had been reported for members in family *Secoviridae*, also surmised for this unassigned species (Carbonell et al. 2016). BRNV infection can be 100% eliminated by Heat treatment and meristem-tip culture for virus-free regenerative plants. (Cheong et al. 2014).

2.6 Plant virus identification and detection tools

Detection of plant viruses and its genetic components are key concepts in identification, characterization and determination of virus type. Detection tools are associated with serological, biological and molecular attributes of viral genes or gene products (Valkonen 2007, Kreuze et al. 2009, Martin et al. 2016). The classical example about Biological Indexing (BI) can be taken from distinguishability of PVY strains. Different strains of PVY are grouped according to serology and infection biology in host species, and plants belonging to the family *Solanaceae* are used as indicator plants. *N. tabacum* could differentiate and elucidate the biological properties of PVY strains (Valkonen 2007). The primary characterized strains of PVY are PVY^{O}, PVY^{C} or PVY^{Z} and PVY^{N}, these are distinguished based on plant defense related hypersensitive response (HR) and extreme resistance (ER) in resistant potato cultivars. PVY^{O}, PVY^{C} or PVY^{Z} are recognized by Ny_{tbr} and Nc_{tbr} genes commonly found in potato resistant cultivars and restricts systemic infection. The Resistance-Breaking (RB) strain PVY^{N} and its new relations are different from others which induces severe venial necrosis in tobacco but occasional HR, limited foliar or no foliar symptoms in resistant potato cultivars. This strain is evolved under recombination in PVY as some strains of PVY^{N} contains PVY^{O} genomic segments and are coined as NTP strains of PVY^{N}. However, it was shown that the general strains PVY^{O} and PVY^{N} difference depends on eight amino-acid residues inside HCPro, and two residues play the key role in recognition of Ny_{tbr}. Thus a structure-function relationship is behind the biological differences exhibited during the infection with PVY in different potato cultivars (Valkonen 2007, Tian et al. 2014). The negativity of this method is of Pollen transmitted *Endornaviruses* that are often symptomless, and latent infection inducing fruit viruses are not graft transmissible for quarantine validations (Martin et al. 2016).

The sensitivity of canonically employed plant virus detection tool, ELISA (enzyme-linked immunosorbent assay) can be explained from above mentioned example of

PVY. PVY^{O} and PVY^{N} strains could be easily detected by serological method of ELISA by either anti-PVY CP mAb (monoclonal antibody) or anti-PVY CP pAb(polyclonal antibody), but efficiency of anti-PVY CP mAb determines relevant identification of PVY strains (Valkonen 2007, Tian et al. 2014). Currently available strain specific anti- PVY CP Abs can detect PVY^{N} specific strains including a variant $PVY^{NTN,}$ but conserved regions of Anti-CP epitopes detect all PVY strains. Anti-PVY mAbs to HCPro epitopes are available to detect PVY^{O} and PVY^{C} strains, also all major strains including PVY^{N} strain (Tian et al. 2014). However, the flip side of the coin is, though every tool availability ensures proper detection, but reliability is added with both biological/molecular and technological upcoming information. For example, a mAb tested and once acted against *Blueberry red ringspot virus* (BRRV) isolate never detected variable isolates from other regions. Three mAb against R15, RBDV resistant-breaking strain reacts with all isolates from red raspberries but only one reacted with an isolate from *Rubus multibracteatus* (Martin et al. 2013).

Polymerase chain reaction (PCR) is widely employed for plant virus detection as nucleic acid assay method (Martin et al. 2016). This is sensitive but straight forward for detection of viral genome by rapid amplification of regions in the genome, either with provided primer information of virus or designing primers from highly conserved domains of related viruses in a family. Almost all plant viral genome sequence deposited in public databases were obtained by PCR amplification of genomes partially or using advanced rapid amplification procedures. Poor primer design or template primer specificity can also result non-specific amplification of plant genetic elements and possibly produce bias in results (Prabha et al. 2013). *Nepovirus* genus and some others of its family *Secoviridae* are highly diverse at nucleotide level, synthesis of oligonucleotide primers possibly cannot detect all available strains, also since these strains are highly adaptive to increase host range. (Martin et al. 2013, Martin et al. 2016). Finally, all these methods may act as barriers for finding unknown unknowns (novel viruses), possibly known unknowns (novel viruses of characterized relatives). The short-time employed microarray technology for pathogen detection also possess limitation above mentioned. Concern in cost of probe design and priori knowledge of nucleotide sequence, at least from related pathogens is requisite for microarrays (Roy et al. 2013). Next Generation Sequencing (NGS) based genome and transcriptome assembly can avail bioinformatic tools to build or rebuild consensus

genome of all organisms. For plant viral diagnosis, its possibility to synchronize input RNAseq/sRNA data information with known virus sequences in online database accentuates rapid detection. Introducing/building algorithms to process and assemble *denovo*, for quick downstream analysis with high sensitivity saves enough task, tariff and time than all abovementioned methods. (Roy et al. 2013, Seguin et al. 2014, Barrero et al. 2017).

2.7 Deep sequencing based virus diagnosis

Deep coverage (X) sequencing of sRNAs, computational analysis and alignment helps to assemble million or billion short sequences known as 'reads'. Generation of lengthy contig (Contiguous set of overlapping sequences) from reads, and further analysis of sequence homology would reveal sRNA biogenesis and target sites in genome/transcriptome (Zheng et al. 2017). Kreuze et al. and Donaire et al. in 2009, first-time showed the complete genome assembly of known and novel plant viruses from vsiRNAs of 21-24nts that are abundant in plants. In the former study done using Ilumina platform, SPFMV and SPCSV were detected from both symptomatic and non-symptomatic sweet potato test plants, along with new viruses belonging to *Mastrevirus* and *Badnavirus* genus. Total vsiRNAs of all size class produced after SPFMV and SPCSV single infections were ~68% unique in contrast to ~53% from plants with Sweet potato virus disease (SPVD) complex. However, 21nt and 22nt vsiRNA size determined the presence of SPFMV and SPCSV from all size class profiles that too belonged to other viruses present or sRNA biogenesis regions in host genome. Three programs were availed to test contig assembly from datasets and Velvet gave faster results, increased accuracy when k-mer (hash length) was set between 13 and 17 (Kreuze et al. 2009). In the later study, examined virus induced targeting of viral genomic regions by vsiRNAs using 454 Pyrosequencing of samples, which were infected with RNA and DNA viruses. The read size composition again supported the predominant 21nt vsiRNA abundance from all RNA viruses except *Cymbidium ring spot virus* (CymRSV), which generated more 22nt than 21nt vsiRNA from genomic regions. Nonetheless, target composition of viral genomes was similar, yet overlapped by 21nt and 22nt total vsiRNAs, unique regional coverage by one size class indeed shared multiple or high density profile of other size class. DCL produced vsiRNAs targeted the same regions with targeting affinities (Donaire et al. 2009).

Both studies confirmed vsiRNA targeting of viral genomes is heterogeneous, meanwhile positive strand of viral RNA is generally mapped well for all viruses detected than the negative strand. Exhaustive vsiRNA sequencing resulted high depth of genomic coverage since the possibility of complete genome assembly relies on maximum possible counts of vsiRNAs (Kreuze et al. 2009, Donaire et al. 2009). Furthermore, target sites indicating hotspot of few viruses were found occurring at highly structured regions promoting DCL activity (Donaire et al. 2009). Complete genomes of viruses belonging to family, *Geminiviridae* (DNA) and *Caulimoviridae* (dsDNA-RT) were constructed following Ilumina sequencing of 19-30nt sRNA profiles. Pervasive nature of vsiRNAs from DNA genomes made the possibility to analyse mutational cloud in viral quasispecies. Red blotch disease of grapevine is known to be the viral complex of DNA viruses, although *Grapevine red blotch-associated virus* (GRBaV) and *Grapevine red leaf-blotch associated virus* (GRLaV) genomes were reconstructed from vsiRNA profiles, *Grapevine yellow speckle viroid I* and *Hop stunt viroid* (HSVd) genomes were additionally assembled from redundant vsiRNAs (Seguin et al. 2013).

Citrus leprosis virus (CiLV) causes indistinguishable symptoms in citrus crops and its low-titre presence in Sweet orange (*Citrus sinensis* L.) was reported. Later studies unveiled the dual nature of this virus in cells, thus two virus types were characterized, CiLV-C (cytoplasmic type) and CiLV-N (nuclear type). Neither ELISA nor RT-PCR revealed a cytoplasmic *Citrus leprosis virus* (CiLV-C2) Type 2 and so sRNA library preparation followed after classical assays. The virus was detected through sRNA omics method after processing reads with bioinformatic assemblers, also subtracting host relative genome derived sRNAs, since no host genome was available. The precise subtraction of redundant and non-redundant reads from two major citrus infecting viroids and one virus narrowed the possibility to cover ~33% RNA-1 and 55% RNA-2 of novel CiLV-C2. This case study pointed the reliability of sRNA omics based detection yet explained possible pitfalls in bioinformatic analysis (Roy et al. 2013). A workflow narrowed for automated viral surveillance and diagnosis (VSD) with bioinformatics tools under YABI platform, pointed the read length criteria of 21, 22, 24nt as significant for detecting viruses of diverse genomes. The read assembly program SPAdes produced longer contigs for ‘denovo’ assembly when k-mer value combination was set (15-17-19-21) and CAP3 scaffold was used (Barrero, et al.

2017).The Universal consideration and availability of NGS method helps to develop online pipelines and novel user-friendly softwares, also for virus diagnosis (Ho & Tzanetakis 2014, Zheng et al. 2017).

The program VirFind was setup for detecting viruses by analysis of RNA-sequencing (RNA-seq) data, indeed provides online server connected service. (Ho & Tzanetakis 2014). VirusDetect is a new software for this type of alignment and assembly for virus detection that is operated with sRNA datasets. This server connected detection tool is designed with automated parameter optimization to build *denovo* genomes based on Velvet assembly (Zheng et al. 2017). Both small as well as large scale sRNA datasets can be analyzed through its inbuilt workflow for getting quick results. Hash length (k-mer) and coverage cut-off are contributing parameters to build contigs and longest contig build would provide positive results. Read redundancy in contigs are removed for homology-dependent mapping, first BLASTN query is performed for best viral sequence matching hits. Further matchless contigs are queried for best hits in BLASTX program. The best result from these queries are mixed to set master genome assembly. This evaluation resulted the discovery of a novel *Potyvirus*, tentatively named *Brazilian weed virus Y* (BWVY) from 9837nt length longest contig assembled. Sanger sequencing confirmed this virus genome assembly and revealed ~1% sequence variation. The outstanding feature in this software is automated generation of graphical outputs (Zheng et al. 2017).

In normal reference guided assembly read alignment through BWA is set. Well aligned reads are assembled to develop contigs that are screened initially in BLASTN and then in BLASTX to revel results. Host genome subtraction is the additional feature set to increase non-redundancy and quality in read profile by removing sufficient sRNAs derived from host genome. Based on the host organism and virus that infects, the databases are categorized, and redundant sequences in certain host databases were removed to improve accuracy. Viruses detected are presented in generated .html file that contain tabulated information of virus name, family and genus, genome coverage, average depth, and no of contigs assembled. Genome coverage by contigs and sequence aligned are generated like histogram graphics in separate .pdf files. Profiling vsiRNA (size) had been shown to detect insect viruses accurately, however it is alerted to confirm the authenticity of viral-contigs generated from such profiles if no homology-dependent assembly is involved (Zheng et al. 2017).

3 OBJECTIVES

Taking into account, the sensitivity and specificity of vsiRNA from sRNA, for virus discovery through deep sequencing, the objective of this study was set as follows:

- Unbiased detection of plant RNA viruses infecting raspberry cultivars in HXR-2 sample pool using VirusDetect.
- Calculating vsiRNA read alignment/mapping with reference genome based on vsiRNA read size.
- RT-PCR detection and identification of RNA viruses present in individual raspberry cultivar samples of HXR-2.
- Coat Protein (CP) encoding nucleotide sequence characterization of virus isolates using molecular cloning-cum-characterization.

4 MATERIALS AND METHODS

4.1 Sampling and methodology

Certified raspberry cultivars openly maintained at MTT Agrifood Research in Piikkiö, Finland were collected by members of 'Rubus germplasm collection' team to evaluate virus persistence in mother plants. The leaf samples from cultivar pairs were harvested, grounded and stored at -80°C.

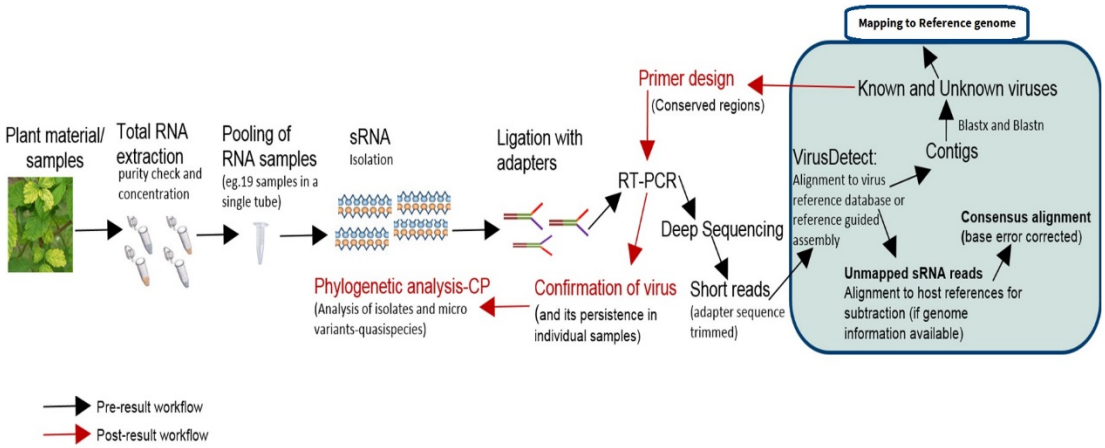


Figure 4. Workflow followed for sRNA deep sequencing based detection and molecular characterization after results.

After detection of viruses using contig assembly program Velvet, soon availability of VirusDetect assisted preparation of workflow shown (Figure 4). For virus identification as post-result part of study, CP/CPs encoding region was selected for primer synthesis, and to characterize RNA virus isolates.

4.2 RNA extraction

Total RNA was extracted following CTAB method from 21 raspberry cultivar leaf samples. Isolation buffer (100ml) was prepared by formulation of 2g CTAB (hexadecyltrimethylammonium bromide), 2g PVP (polyvinylpyrrolidone), 5ml 0.5M EDTA, 10ml 1M Tris-HCL and 40ml 5M Nacl. The isolation buffer was heated to 65°C in water bath, and 200µl β-mercaptoethanol was added to 10 ml of isolation buffer to obtain final concentration of 2%. 1ml of isolation buffer was added to 0.15 g ground leaf samples collected in 2ml collection tubes. 1ml Chloroform/Isoamyl alcohol (24:1) was added to tubes for 15 min strong homogenization and centrifuged at 10000g for 15 min at 4°C. Supernatant was transferred to new collection tubes and equal amount of Chloroform/Isoamyl alcohol (24:1) homogenization step was

repeated, followed through centrifugation at 10000g for 15 min at 4°C. 10M Lithium chloride was added to supernatant collected measuring ¼ volume and mixed gently. The mixture was kept at 4°C overnight for RNA precipitation. 500µl 70% ice-cold ethanol was added into precipitation mixture tubes and centrifuged at 10000 g for 20 min at 4°C. The pellet was carefully dissolved in 500µl warmed SSTE buffer (2ml 5M Nacl, 500µl 10% SDS, 100µl 1M Tris-HCL and 20µl 0.5M EDTA. 1ml of Chloroform/Isoamyl alcohol (24:1) was added and centrifuged at 10000 g for 15 min at 4°C. The clear lysate was transferred to 1.5ml Eppendorf tube following addition of 2 volume absolute ethanol. The Precipitation was kept for 2 hours at -80°C and centrifuged at 13000 g for 20 min at 4°C. The pellet was washed with 70% ethanol and centrifuged twice at 13000 g for 1 min at 4°C. The tubes were carefully dried and RNA pellet was dissolved in 40µl autoclaved Milli-Q water. The concentration and purity (A260/A280) of the total RNA extracted was measured using Nano drop spectrophotometer UV–vis (ThermoFisher Scientific, USA) and stored at -80°C.

4.3 sRNA deep sequencing

Sample pool HXR-2 was prepared by the collection team and sent to NGS sequencing facility, Fasteris SA (Plan-les-Outes, Switzerland). High-throughput small RNA sRNA) sequencing was performed as reported (Nordenstedt et al. 2017) except the method- DNA single-end (unpaired) sequencing. Pre-processed quality reads of 21-24nt fastq. file was obtained for downstream analysis.

4.4 Genome alignment and computation

4.4.1 VirusDetect

Automated VirusDetect was employed for rapid detection of viruses and viroid for this study. VirusDetect is available in Chipster of CSC Finland (Kallio et al. 2011), default parameters are automatically set for single click run. The pre-processed, HXR-2.fastq file (21-24nt) was loaded into VirusDetect and job session was executed following default parameters. The default parameter described - Minimum fraction of a contig covered by reference- 0.75 (i.e. 75% of assembled contig length should match with reference genome), Minimum fraction of virus reference covered by contigs- 0.1 (i.e. 10% of the reference must match a contig build), Minimum read depth- 5 (a nucleotide in reference should cover at least 5 times for consideration coverage) The .fastq files

were sorted for sub-class size by filtering 21, 22, 23 and 24nt read size separately using PRINSEQ program in Chipster.

4.4.2 Genome mapping/coverage computation

Bowtie2 short-read aligner was used for aligning vsiRNA reads of all (nt) size class (21, 22, 23, 24) and single size sub class to reference genomes (Langmead & Salzberg 2012) After bowtie2-build alignment of reads, the resulting output in SAM (Sequence Alignment Map) format was converted into BAM (Binary Alignment Map) format availing SAM tool commands- samtools view, also sorted- samtools sort. (Li et al. 2009). The sorted file was converted to a text file, extracting read no. using BED tools command bedtools genomecov. (Quinlan & Hall 2010). The read values extracted in text file were analyzed in MS Excel and histograms were plotted after calculation. Computational operations were performed connecting to taito terminal (Taito terminal. 2013) of CSC Finland through putty (windows) command line. WinSCP transfer is used for transferring files.

4.5 Reverse-Transcription Polymerase Chain Reaction (RT-PCR)

Total RNA isolated from individual cultivars of HXR-2 were subjected for first strand cDNA synthesis. In Reverse Transcription (RT) reaction tubes/reaction, 1µl Random primers (200µg/ml) and 1µl total RNA from samples in addition to 10.5µl Milli-Q was added and heated at 65°C for 5 minutes, and placed immediately on ice for 5 minutes. Then, 4µl 5x RT Reaction buffer (ThermoFisher, Lithuania), 2µl dNTP (10mM) mix, 0.5µl (20units/µl) Thermo Scientific Ribolock RNase (ThermoFisher, Lithuania) and 1µl(200 units/µl) Revert Aid Reverse Transcriptase (ThermoFisher, Lithuania) was added to tubes first before gentle mixing, and incubated at room temperature for 10 minutes, followed by incubation at 42°C for 60 more minutes. The reaction was stopped by incubating 70°C for 10 minutes, and transferred to ice, then stored at -20°C. Polymerase Chain Reaction (RT-PCR) was executed based on protocols mentioned in DyNAzyme II DNA polymerase (ThermoFisher Lithuania), Phusion HF DNA Polymerase (ThermoFisher, Lithuania) and Taq DNA polymerase (ThermoFisher, WI, USA) kits, reaction components and primers used are described (Table 1, Table 2, Table 3, Table 4, Table 5, Table 6).

Table 1. PCR reaction components for RBDV detection of 245bp CP region.

PCR reaction components			PCR cycles
Master mix	10X DyNAzyme buffer	2µl	
	dNTPs 10 mM	0.5µl	Initial Denaturation- 94°C 2 min
	$MgCl_2$	0.2µl	Denaturation - 94°C 30 sec
	5'-3'- U2F	0.4µl	Annealing(Primer) - 57.5°C 30 sec x35
	3'-5'- L3R	0.4µl	Extension - 72°C 30 sec
	DyNAzyme II DNA polymerase	0.2µl	Final extension - 72°C 8 min
	Water	14.3µl	Cooling 4°C
Template: cDNA		2µl	
Total		20µl	

Table 2. PCR reaction components for RBDV detection of 872bp CP region.

PCR reaction components			PCR cycles
Master mix	10X DyNAzyme buffer	2µl	
	dNTPs 10 mM	0.5µl	Initial Denaturation- 94°C 2 min
	$MgCl_2$	0.2µl	Denaturation - 94°C 30 sec
	5'-3'- U1	0.4µl	Annealing(Primer) -55.3°C 30 sec x35
	3'-5'- L4	0.4µl	Extension - 72°C 40 sec
	DyNAzyme II DNA polymerase	0.2µl	Final extension - 72°C 8 min
	Water	14.3µl	Cooling 4°C
Template: cDNA		2µl	
Total		20µl	

Table 3. PCR reaction components for BRNV detection of 155bp CPs region.

PCR reaction components			PCR cycles
Master mix	10X DyNAzyme buffer	2µl	
	dNTPs 10 mM	0.5µl	Initial denaturation - 94°C 2 min
	$MgCl_2$	0.2µl	Denaturation - 94°C 30 sec
	5'-3'- CPpieni1F	0.4µl	Annealing(Primer) - 53.3°C 30 sec x35
	3'-5'- Small2R	0.4µl	Extension - 72°C 30 sec
	DyNAzyme II DNA polymerase	0.2µl	Final extension - 72°C 8 min
	Water	14.3µl	Cooling 4°C
Template: cDNA		2µl	
Total		20µl	

Table 4. PCR reaction components for BRNV detection/sequencing of 1459bp CPs region.

PCR reaction components			PCR cycles
Master mix	5X Phusion HF Buffer	4µl	Initial denaturation - 98°C 1 min Denaturation - 98°C 10 sec Annealing(Primer) -54.2°C 20 sec x35 Extension - 72°C 65 sec Final extension - 72°C 8 min Cooling 4°C
	dNTPs 10 mM	0.5µl	
	$MgCl_2$	0.2µl	
	5'-3'- F_1BRNV	1µl	
	3'-5'- R_2BRNV	1µl	
	Phusion HF DNA polymerase	0.2µl	
	Water	11.1µl	
Template: cDNA		2µl	
Total		20µl	

Table 5. Reaction components for colony PCR screening of positive clones

PCR reaction components			PCR cycles
Master mix	10X Taq DNA polymerase buffer	2µl	Initial denaturation -95°C 2 min Denaturation - 95°C 30 sec Annealing(Primer) - 47°C 30 sec x30 Extension - 72°C 80 sec Final extension - 72°C 8 min Cooling 4°C
	dNTPs 10 mM	0.5µl	
	$MgCl_2$	2µl	
	5'-3' *	0.5µl	
	3'-5' *	0.5µl	
	Taq DNA polymerase	0.1µl	
	Water	14.4µl	
Total		20µl	* Primer of interest

Table 6. Primer pair used for detection and Sanger sequencing of RBDV and BRNV.

Primer	5'-3'	Length (nt)	Reference
RBDV			
U2	TTCATCCTCCAAATCTCAGCAAC		
L3	CGTCGACGGCACCGCCCACCACA	245	Kokko et.al(1996)
U1	GCTGTTCCACCAATCGTTA		
L4	GCTATGCCGTTTATCTCAC	872	Kokko et.al(1996)
BRNV			
CP pieni1F	CACWCCNGGAMAGGACTA		3880-4035 CPs
Small2R	TTCTCTGGGCGCAAATCTAT	155	(DQ344640)
F_1BRNV	TATCTCGTACTCCACCAAC		9- 1467 CPs
R_2BRNV	CTATAGACCTAGGGCACC	1459	(DQ344640)

4.6 Molecular cloning into pGEM®-T easy vector

Molecular cloning formed the elementary work of genetic engineering and was done using pGEM®-T easy vector (Promega, WI, USA), protocol and comprised of following, also additional steps.

Purification: The PCR product was gel excised and dissolved in binding buffer following the protocol mentioned in E.Z.N.A® Gel Extraction kit (Omega BIO-TEK, Norcross, GA, USA).

A-tailing: The RT-PCR product of BRNV (1459nt) was amplified using Phusion HF polymerase. Poly-A tailing (TA cloning) based on Poly(A) tailing kit (ThermoFisher Scientific, Lithuania) was done to the purified product (7µl) by adding 0.2mM dATP (1µl), 1x optimized buffer (1µl) and 1U DyNAzyme II DNA polymerase (1µl) followed by incubation at 72°C for 20 min.

Ligation: T4 DNA Ligase was used for ligation of insert into pGEM®-T easy vector (3:1 ratio) according to the protocol (Promega, WI, USA).

Transformation: Laboratory prepared *E.coli* (DH5α) competent cells in tubes were initially thawed in ice, and 5µl of ligation mixture was added. The tube contents were mixed well by flicking and left on ice for 20 min at room temperature. Transformation was done at 42°C for 90 seconds, immediately transferred to ice. 500µl of fresh LB medium was added to incubate transformants at 37°C for 60 min. The transformation mixture was mixed by pipetting and streaked to plates- LB + Ampicillin (100mg/ml) + X-Gal (5-bromo-4-chloro-3-indolyl-β-D galactopyranoside) + 1M IPTG (Isopropyl β-D-1-thiogalactopyranoside).

Restriction digestion: Sanger sequenced and labeled BRNV isolates were subjected to restriction digestion endonuclease activity of *Pst*I (CTGCA^G site) and *Eco*RI (G^AATTC) according to Restriction Enzyme Cloning (ThermoFisher Scientific, USA). 10µl of Plasmid was incubated for 60min at 37°C.

4.7 Agarose-gel electrophoresis

All RT-PCR amplified products were subjected to Gel electrophoresis. Agarose Gel of 1.5% was prepared by heating 4.5g agarose in an Erlenmeyer flask containing 300ml 1xTAE buffer (242g Tris + 57.1 glacial acetic acid + 100ml 0.5M EDTA (pH

8.0), 50X buffer per litre), and 20µl PCR product was loaded in gel, to check length of amplified fragments of RBDV (245bp) and BRNV (155bp), running at 90-140V along with a DNA ladder of 1Kb (ThermoFisher Scientific, Lithuania). Amplified 20µl PCR products of 872bp RBDV CP, 1459bp BRNV CPs, and Colony PCR and pGEM ®-T easy vector restriction digestion (10µl) were analysed in 1% agarose gel electrophoresis with Gene ruler ladder mix (ThermoFisher Scientific, Lithuania). The gel purified products were subjected to Sanger sequencing at Macrogen unit in Europe (Amsterdam, The Netherlands).

4.8 Phylogenetic analysis

Neighbor-joining tree construction was performed in MEGA7 (Molecular evolutionary genetics analysis). The multiple sequence alignment by ClustalW was first completed in MEGA7 and exported .mega file was loaded into Neighbor-joining tree construction algorithm through nucleotide substitution type of Kimura-2 by bootstrapping 100 replicates, gamma distribution. The missing data in alignment was selected for pairwise deletion (Kumar et al. 2016). Pairwise sequence identity alignment was calculated for both nucleotide and protein sequences of accessions. Alignment was done in MEGA 7, multi-aligned sequence .fasta file was uploaded in SDT program (windows) and single-click execution of pairwise calculation was done with default parameters (Muhire et al. 2014). BLAST (Basic Local Alignment tool) was used throughout all sequence analysis for finding homology between sequences in NCBI (Altschup et al. 1990). Multialign interface page was used extensively checking alignment for sequenced products (Corpet 1998). *NEB* cutter online was used to cleave DNA sequences of isolates with endonuclease restriction enzymes (Vincze et al. 2003).

5 RESULTS

5.1 Viruses detected by VirusDetect:

VirusDetect reaffirmed the presence of two major *Rubus* infecting (+) RNA viruses: *Raspberry bushy dwarf virus* (RBDV) and *Black raspberry necrosis virus* (BRNV), in HXR-2 sample pool. Qualitatively pre-processed data file containing 21-24nt vsiRNA reads aligned to reference that also assembled contigs for virus detection (Figure 5). Since the objective of study was focused on RNA viruses of raspberry, other viruses detected were excluded from results. RBDV RNA-1(KJ007639) was covered 98% by 11 contigs, and longest contig length was 1951nt that masked the 5'end of genome. The longest contig assembled across RNA-2 (KJ007640) was 673nt, whole genome was 93.5 % covered by 10 contigs (Table 7). Contig assembly consensus marked with red bar showed adequate read alignment to RBDV reference genomes (Figure 5(2)).

BRNV full length genomes of three diverse isolates aligned with assembled contigs. The longest contig assembly to all BRNV genomes resulted 746-734nt length at the 3'end of RNA-1 and RNA-2 (Table 7, Figure 5(1)). North American isolate of BRNV RNA-2 (DQ344640) was uniquely covered (84.1%) by 53 contigs, meanwhile RNA-1 of all other BRNV isolates were scarcely covered (30%) by 15-25 contigs. RNA-2 of BRNV-Alyth was covered (42.9%) by 34 contigs, yet BRNV-FI coverage (27%) was reduced to 23 contigs (Table 7). The protein en-coding nts(s) regions in RNA-1 and RNA-2 were barely aligned with contigs, besides one narrow part in large Coat Protein (CP1) region where 2 contigs aligned (Figure 5(1)). Read alignment and contig assembly to BRNV-NA RNA-1 (DQ344639) and RNA-2 (DQ344640) pointed excellent read count availability at 3'UTR of genome. This showed that reference coverage (%) calculation in VirusDetect was based on contig number and length. No host sequences were available for host genome subtraction, yet 21-24nt vsiRNA read alignment and contig assembly showed considerable coverage to RBDV genomes, and poor mapping in BRNV isolates. Average depth value of vsiRNA and percentage whole genome coverage (Table 7) shown for viruses were not correlative since vsiRNA ubiquitous targeting genomes had been reported earlier in sRNA profiling studies. VirusDetect had not detected above mentioned viruses or other viruses, when sorted sRNA reads fastq. file of unique 24nt size was analysed for results. Results collectively obtained here endorsed vsiRNA downstream analysis.

Table 7. Complete genome coverage information of viruses detected by VirusDetect in sample pool HXR-2

Virus	Reference genome	Length	Reference coverage (%) across genome	Average depth	Number of contigs	Longest contig length
RBDV	KJ007639 (RNA-1)	5449	98.0 (5338)	193.2	11	1951
	KJ007640 (RNA-2)	2231	93.5 (2085)	97.4	10	673
BRNV-NA	DQ344639 (RNA-1)	7581	30.0 (2277)	131.1	24	734
	DQ344640 (RNA-2)	6364	84.1 (5354)	70.3	53	742
BRNV- Alyth	FN908128 (RNA-1)	7572	24.4 (1851)	142.2	20	741
	FN908129 (RNA-2)	6350	42.9 (2721)	140.5	34	741
BRNV- FI	HE611022 (RNA-1)	7528	21.2 (1597)	147.7	15	741
	HE614901 (RNA-2)	6326	27.0 (1707)	136.8	23	746

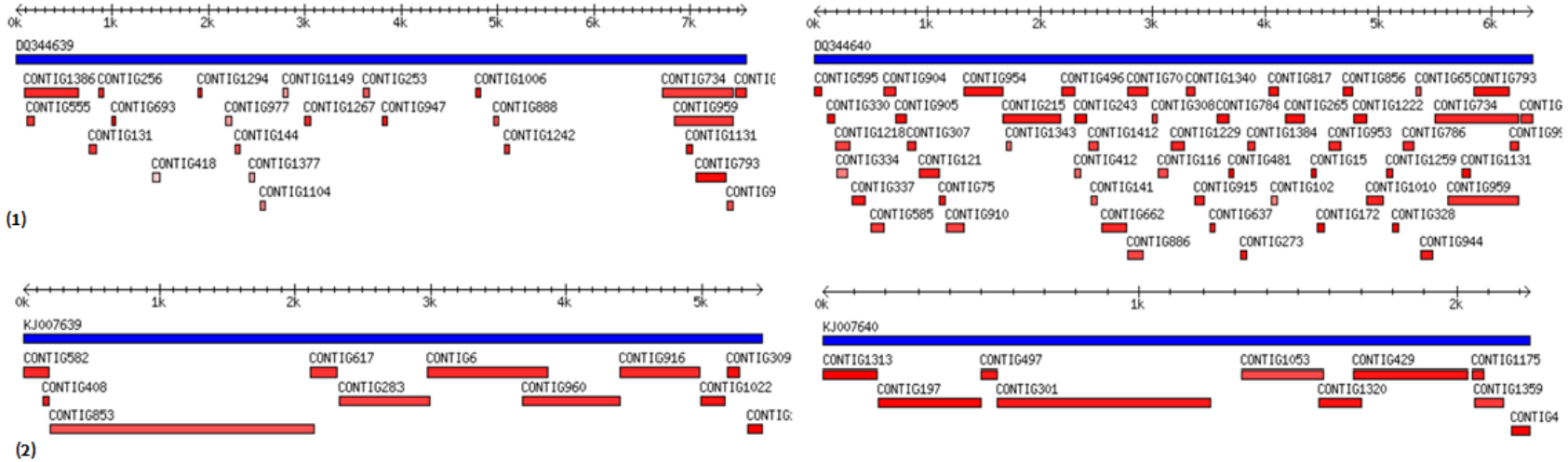


Figure 5. Contigs assembled by VirusDetect aligning across the reference genome of **(1)** BRNV and **(2)** RBDV. Accession numbers are shown.

5.2 Genome mapping/coverage by vsiRNAs

RBDV (KJ007639, KJ007640) and BRNV-NA (DQ344639, DQ344640) accessions were subjected for genome alignment and mapping analysis based on vsiRNA read size. Overall sRNA profile computation and vsiRNA read alignment indicated the significance of read size, since 21nt vsiRNA reads were calculated preponderant, followed by 22nt, and 24nt vsiRNA reads inadequately aligned with reference genomes (Table 8). RBDV genomes were well aligned with 21nt and 22nt vsiRNA reads, 23 and 24nt vsiRNA reads were calculated redundant in reference alignment. BRNV-NA genomes adequately aligned with 21nt and 22nt reads, those also positioned as cardinal vsiRNA size in contrast to other size reads (Table 8).

Table 8. Reference genome alignment by vsiRNA read class size

Viral RNA	Reference ID	Length	Number of reads aligned with reference				
			All size reads/ total	21nt	22nt	23nt	24nt
RBDV RNA-1	KJ007639	5449	32643	17884	12959	319	319
RBDV RNA-2	KJ007640	2231	9363	5576	3343	331	113
BRNV RNA-1	DQ344639	7581	8677	4286	3752	374	265
BRNV RNA-2	DQ344640	6364	9258	4484	4074	434	266

Reads aligned to reference might reflect the total vsiRNA read count or average depth, but vsiRNAs mapping reference was checked separately to indicate hotspots across whole genome. 21 and 22nt vsiRNA reads were mapped to both viral genomes that also spanned across full length (Figure 6, Figure 7). Genomic regions covered by vsiRNAs of 21nt and 22nt read size was observed significant to calculate least covered and most covered region (Table 9, and Table 10).

RBDV genomes mapped from 5’end showed clear targeting by vsiRNAs, and reduced targeting towards the 3’end. In both RNA molecules, vsiRNA map at 3’UTR nucleotides exhibited scarce coverage, it was moreover observed as the least covered genomic region (Figure 6, Table 9). This was contrary to the mapping from 5’ UTR towards the coding regions, which was distinguishable in RNA-1 (Figure 6). The hotspot in RBDV RNA-1 occurred at 992nt 5’ Replicase region, and this nucleotide position (indicated with blue arrow) was covered mostly covered by vsiRNAs of 21nt size. One hot spot at 327nt in RNA-2 MP encoding region was spotted (blue arrow), also mostly covered by 21nt read size in RNA-2 (Figure. 6).

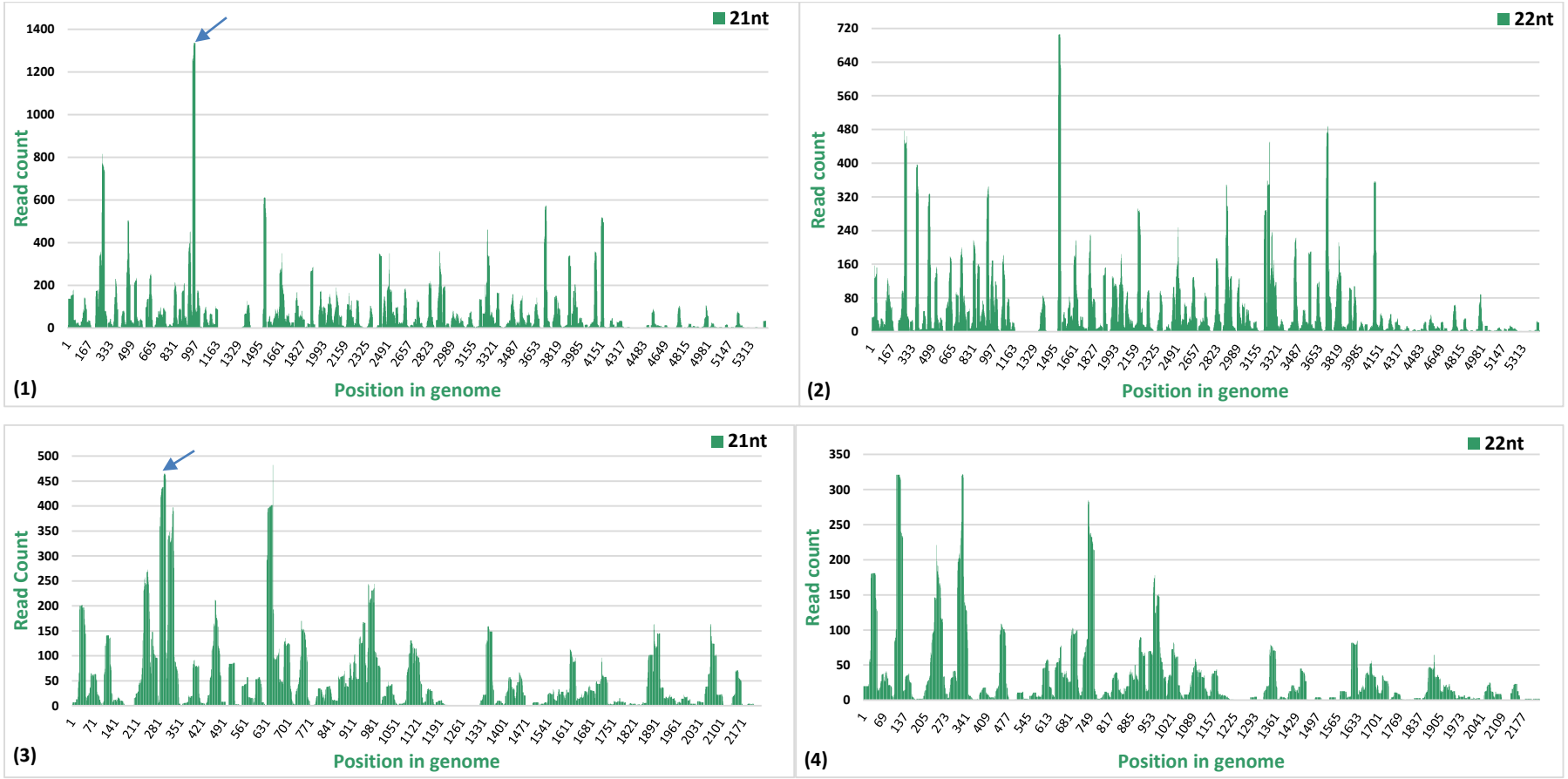


Figure 6. vsiRNAs mapped to each aligned nucleotide position in RBDV genome segregated by vsiRNA read size -21nt & 22nt. **(1)** RNA-1(KJ007639) -21nt **(2)** RNA-1(KJ007639) -22nt **(3)** RNA-2 (KJ007640) -21nt **(4)** RNA-2(KJ007640) -22nt. The X-axis scale is Position in genome and Y-axis scale is Read count. The hotspot in 21nt read size is marked using blue arrow.

Table 9. vsiRNA reads covering the regions in RBDV RNA-1 and RBDV RNA-2

Reference ID/ Reference region * Least covered region is marked	**Length/ nt**	**Number of nt(s) in genomic regions covered by unique size reads**			
		21nt	**22nt**	**23nt**	**24nt**
RBDV-RNA-1 (KJ007639)	5449	2749	2513	936	320
5'UTR	103	56	46	25	20
Replicase	5085	2636	2425	897	298
P12(putative)	330	80	56	13	2
3'UTR*	53	9	12	1	1
RBDV-RNA-2 (KJ007640)	2231	1035	893	311	107
5'UTR	131	78	65	27	15
MP	1077	598	564	208	58
CP	825	335	249	71	29
3'UTR*	81	19	7	5	4

5'UTR and 3'UTR in BRNV genomes showed clear vsiRNA coverage. The near complete missed region in RNA-1(DQ344639) was RdRp, and in RNA-2 (DQ344640) vsiRNAs unambiguously not mapped the CPs encoding region (Table 10). The histogram map showed an increase of read count number (depth) in nucleotides succeeding RdRp encoding region, a similar pattern was spotted after CPs encoding nucleotides (Figure 7). In RNA-1 vsiRNA hotspot occurred at 6851nt position, at 5637nt position in RNA-2, both are 3'UTR regions of genome. The read size of vsiRNA 22nt makes this depth of coverage at positions (blue arrow), and similar pattern of vsiRNA targeting the whole 3'UTR region in BRNV genomes was clearly noted (Figure 7).

23nt and 24nt vsiRNA read size coverage was comparatively low and the necessity in genome mapping was found to be non-significant of its low mapping to reference genomes (histograms not shown). The sudden decrease in read count of 23 and 24nt vsiRNAs was calculated. Both vsiRNA read alignment and coverage computation of genomic regions discriminated the efficiency of read size, and it is possible to deduce that 23nt and 24nt vsiRNAs covered the same positions of 21nt and 22nt vsiRNAs (Table 8, Table 9, Table 10).

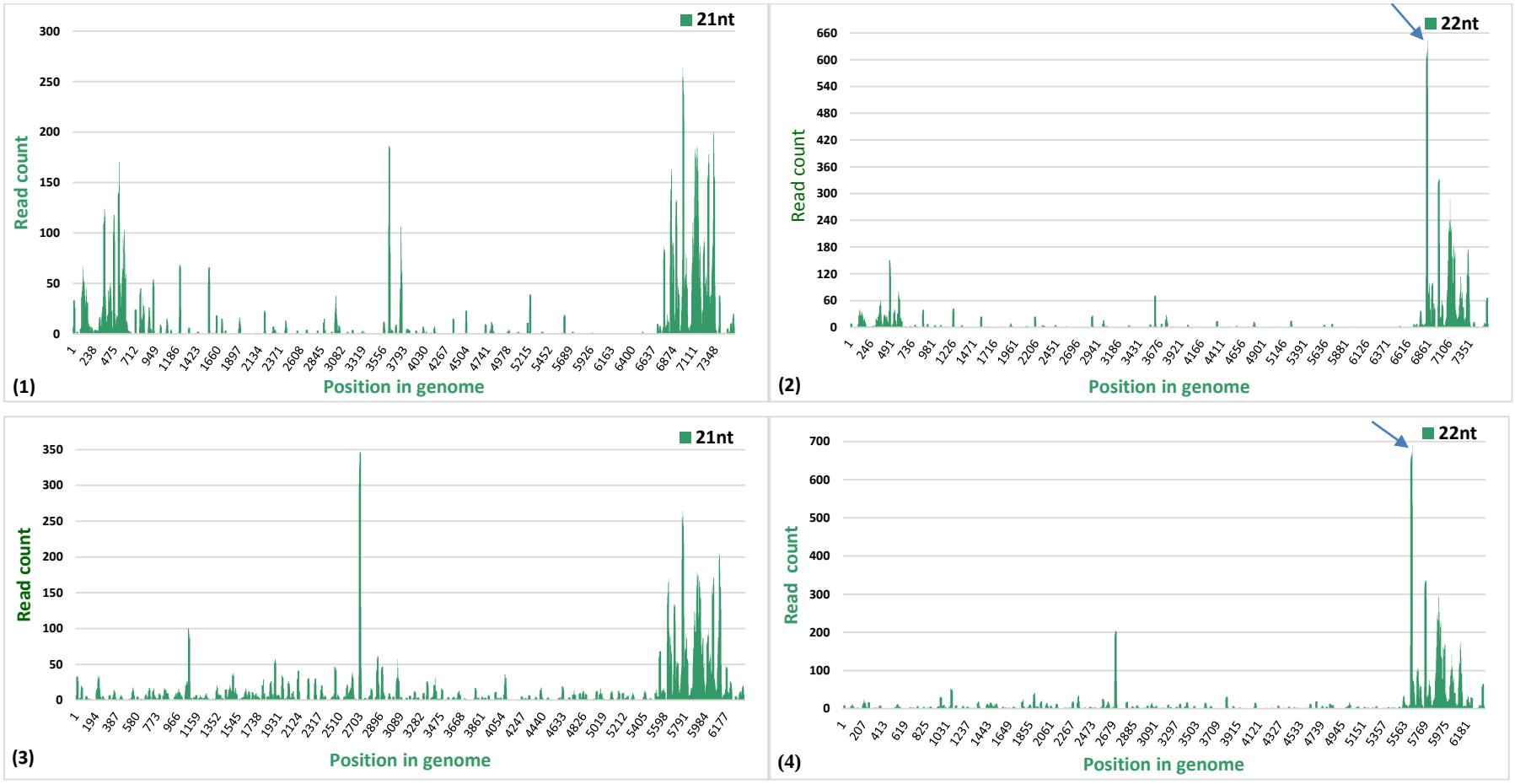


Figure 7. vsiRNAs mapped to each aligned nucleotide position in BRNV genome segregated by vsiRNA read size -21nt & 22nt. **(1)** RNA-1(DQ344639) -21nt **(2)** RNA-1(DQ344639) -22nt **(3)** RNA-2(DQ344640) -21nt **(4)** RNA-2(DQ344640) -22nt. The X-axis scale is Position in genome and Y-axis scale is Read count. The hotspot in 22nt read size is marked using blue arrow.

Table 10. vsiRNA reads covering regions in BRNV RNA-1 and BRNV RNA-2

Reference ID/ Reference region * Least covered region is marked	**Length/ nt**	**Number of nt(s) in genomic regions covered by unique size reads**			
		21nt	**22nt**	**23nt**	**24nt**
BRNV RNA-1 (DQ344639)	7581	1184	884	269	103
5'UTR	146	45	34	11	18
ProC	1545	399	279	68	18
Helicase	1515	116	49	3	2
VPg	78	3	1	0	0
Protease	702	79	50	5	3
RdRp*	2652	86	48	9	0
3'UTR	943	456	423	173	62
BRNV RNA-2 (DQ344640)	6364	1690	1295	318	102
5'UTR	223	43	41	9	13
MP	972	223	148	27	6
CP1	2763	682	491	78	20
CPs*	1467	226	157	14	2
3'UTR	939	506	458	190	61

5.3 Virus detection by RT-PCR

Plant RNA viruses detected by VirusDetect were confirmed using RT-PCR from individual raspberry samples in HXR-2 pool. For RBDV Coat Protein (CP) encoding region was targeted, and BRNV small Coat Protein (CPs) encoding was selected. RBDV was detected in 5 samples by amplification of 245bp CP region, and BRNV presence was confirmed in 6 samples following amplification and Sanger sequencing of 155bp or 1459bp CPs region (Figure 8, Figure 9). Complete 872nt RBDV CP encoding region (sgRNA) recovered from cultivar 'Ville' (38) was amplified through RT-PCR, and introduced into pGEM®-T easy vector, positive clones were sequenced for downstream phylogenetic analysis. From 4 cultivars, partial BRNV CPs region RT-PCR preceded 1459bp, which was used for vector cloning of isolates (Figure 6). The primer sequences used for RT-PCR is listed (Table 6). RBDV was detected in

raspberry samples: RU25 Norna, Ranta Kaukonen and Ville, and from RU158 Hoolin Kanta, RU24 Heija and RU18 Heisa BRNV CPs sequence information was obtained. Confirmation of RBDV and BRNV positives were observed in same cultivar (pair) of HXR-2 sample pool, and labeled BRNV isolates sequence were deposited in EMBL nucleotide database, accession no. shown (Table 11). No mixed infection among RNA viruses was detected in raspberry samples throughout RT-PCR assay, despite laborious detection for BRNV in samples were performed. BRNV amplification followed gel electrophoresis which developed just smear in gel indicated its recalcitrant nature to characterization (Figure 9).

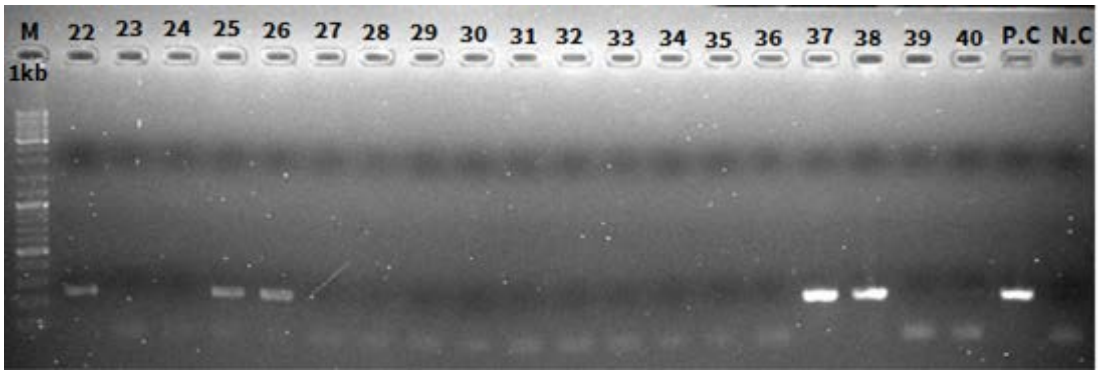


Figure 8. Agarose gel electrophoresis confirmation of RBDV in individual raspberry cultivars of sample pool HXR-2. 245bp CP region of RNA-2 was amplified from positive samples. Sample lanes are marked 22-40 from left to right. Marker (M), positive (P.C) and negative controls (N.C) are indicated.

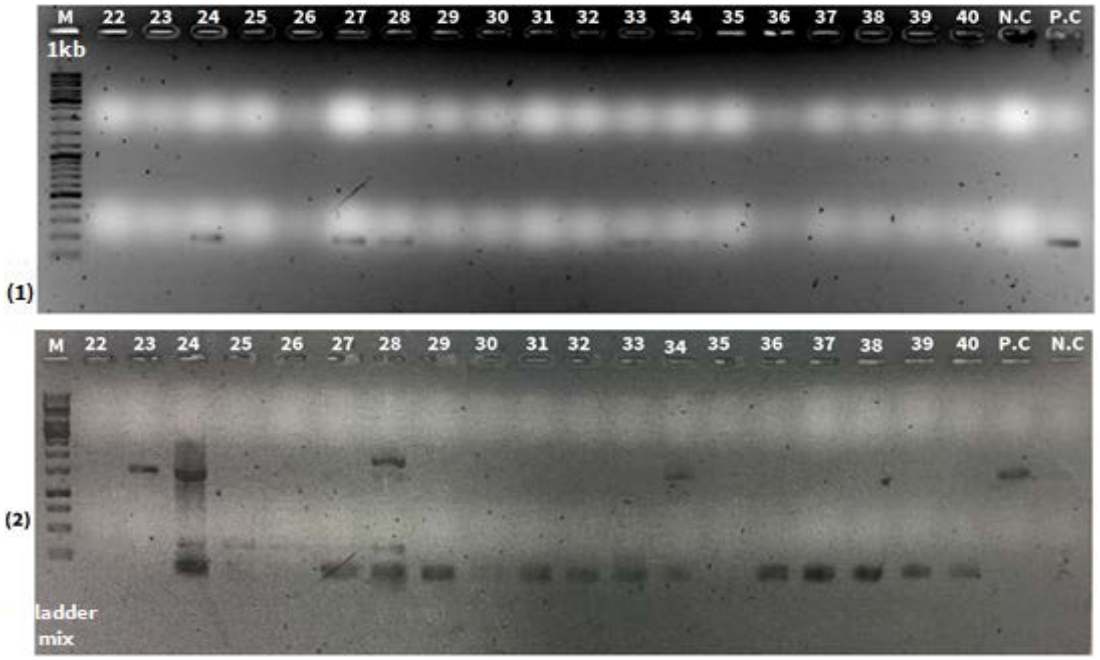


Figure 9. Agarose gel electrophoresis confirmation of BRNV in individual raspberry cultivars of sample pool HXR-2. CPs regions of RNA-2 **(1)** 155bp or **(2)** 1459bp were amplified from virus positive samples. Samples lanes are marked 22-40 from left to right. Marker (M), positive (P.C) and negative controls (N.C) used are indicated.

Table 11. RT-PCR results -Virus positives in raspberry cultivars of sample pool HXR-2

Sample no.	Cultivar	RBDV	BRNV		
		U2/L3	CPpieni1F /Small2R	F_1BRNV/R_2BRNV [+] Accession no. is indicated	
22	RU25 Norna	+	-	-	
23	RU158 Hoolin Kanta	-	-	+	FI:HOOH23:15
24	RU158 Hoolin Kanta	-	+	+	FI:HOOH24:15
25	Ranta, kaukonen	+	-	-	
26	Ranta, kaukonen	+	-	-	
27	RU24, Heija	-	+	-	
28	RU24, Heija	-	+	+	FI:HEIJ28:15
29	Ojanperä, Kaukonen	-	-	-	
30	Ojanperä, Kaukonen	-	-	-	
31	R. all. Mäjestät	-	-	-	
32	R. all. Mäjestät	-	-	-	
33	RU18 Heisa	-	+		
34	RU18 Heisa	-	+	+	FI:HEIS34:15
35	HY 71029	-	-	-	
36	HY 71029	-	-	-	
37	Ville	+	-	-	
38	Ville	+	-	-	
39	Indian Summer	-	-	-	
40	Indian Summer	-	-		

5.4 Molecular characterization of CP encoding region

CP encoding sequences from 12 RBDV isolates were aligned to test neighbour joining tree phylogeny, along with the newly sequenced 872bp (nt) obtained from cultivar Ville-38, marked RBDV38. The phylogeny mostly distinguished isolates based on host plant species rather than geographical origin (Figure 10). The clear distinct isolate from *R multibracteatus* (China) grouped outside, and was found diverse (87% identity) at nucleotide level. The isolate RBDV38 showed precise nucleotide identity (99.9 %) with an isolate (AY894678.1) previously sequenced from Finland, these isolates formed a separate clade (100% bootstrap) in the branch that included red raspberry RBDV isolates. RBDV38 isolate cluster included Scottish isolate (D200), the one widely considered for RBDV serological studies, but D200 aligned in the clade of Canadian isolate with 100% boot strap support. All these isolates are identical (1% difference) at nucleotide level. Isolates from Sweden and UK formed a separate clade, and showed 97-99% nt identity (100% bootstrap). Grapevine (*Vitis vinifera* L.) isolate revealed 94-95% nt identity with all red raspberry isolates, but formed a separate branch in tree Phylogeny. All raspberry isolates within exhibited >96% identity, however Ecuadoran (KC315892.1) isolate from Andean blackberry (*Rubus glaucus* Benth.) showed 95% identity with raspberry isolates (Figure 10).

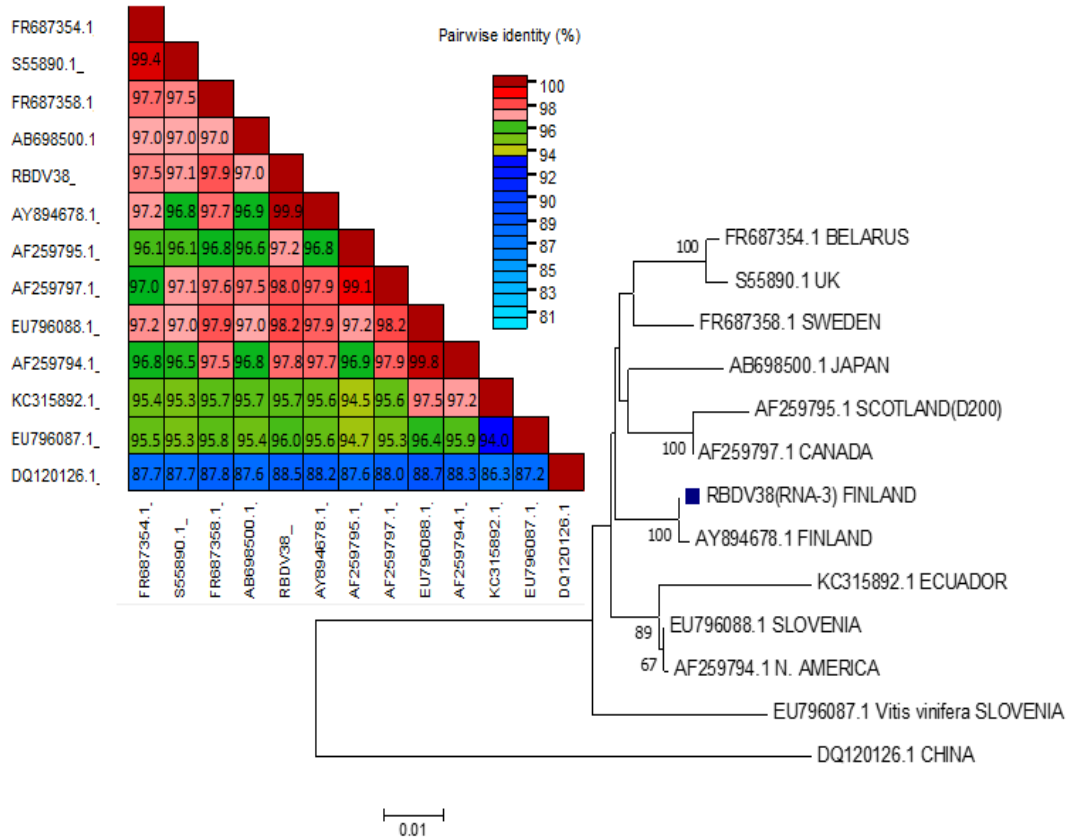


Figure 10. (1) Unrooted neighbour-joining phylogenetic analysis of Coat Protein (CP)-encoding nucleotide (nt) sequences from 13 RBDV isolates. Accession numbers and boot strap values >50 are shown. Kimura 2-parameter model, Gamma distributed (G). RBDV isolate (38) sequenced from Finland is marked with blue rectangle and RBDV isolate from grapevine is indicated. **(2)** Pairwise comparison of RBDV CP-encoding (nt) accessions –**lower diagonal left**. The colour reference represents pairwise nucleotide identities between corresponding isolates.

BRNV CPs encoding nucleotide accessions from Finland, UK (Alyth) and North America (NA) were grouped into four different branches. The sequenced isolates from raspberry cultivars in Finland are FI:HOOH23:15, FI:HOOH24:15, FI:HEIJ28:15 and FI:HEIS34:15, these clustered with isolates from wild raspberries. The isolate from cultivar (RU24 Heija) FI:HEIJ28:15 aligned in the clade formed along 13 BRNV isolates from wild raspberries. It showed 99.9% identity at nucleotide (nt) level and 100% identity at amino acid (aa) level with FI:HAU29A:09 in the same clade. Other BRNV isolates from raspberry cultivars in Finland were grouped far with FI:VII15:09 and FI:VII17:09. These isolates were 96-99% (nt) and 98-99% (aa) identical inside the group. FI:VII17:09 identity matched 98.7% (nt) and 99.1% (aa) with FI:HEIS34:15, the isolate sequence obtained from cultivar RU18 Heisa. FI:HOOH23:15 and FI:HOOH24:15 are isolates recovered from cultivar Hoolin Kanta, showed 99.7% (nt) and 99.6% (aa) identity (Figure 11).

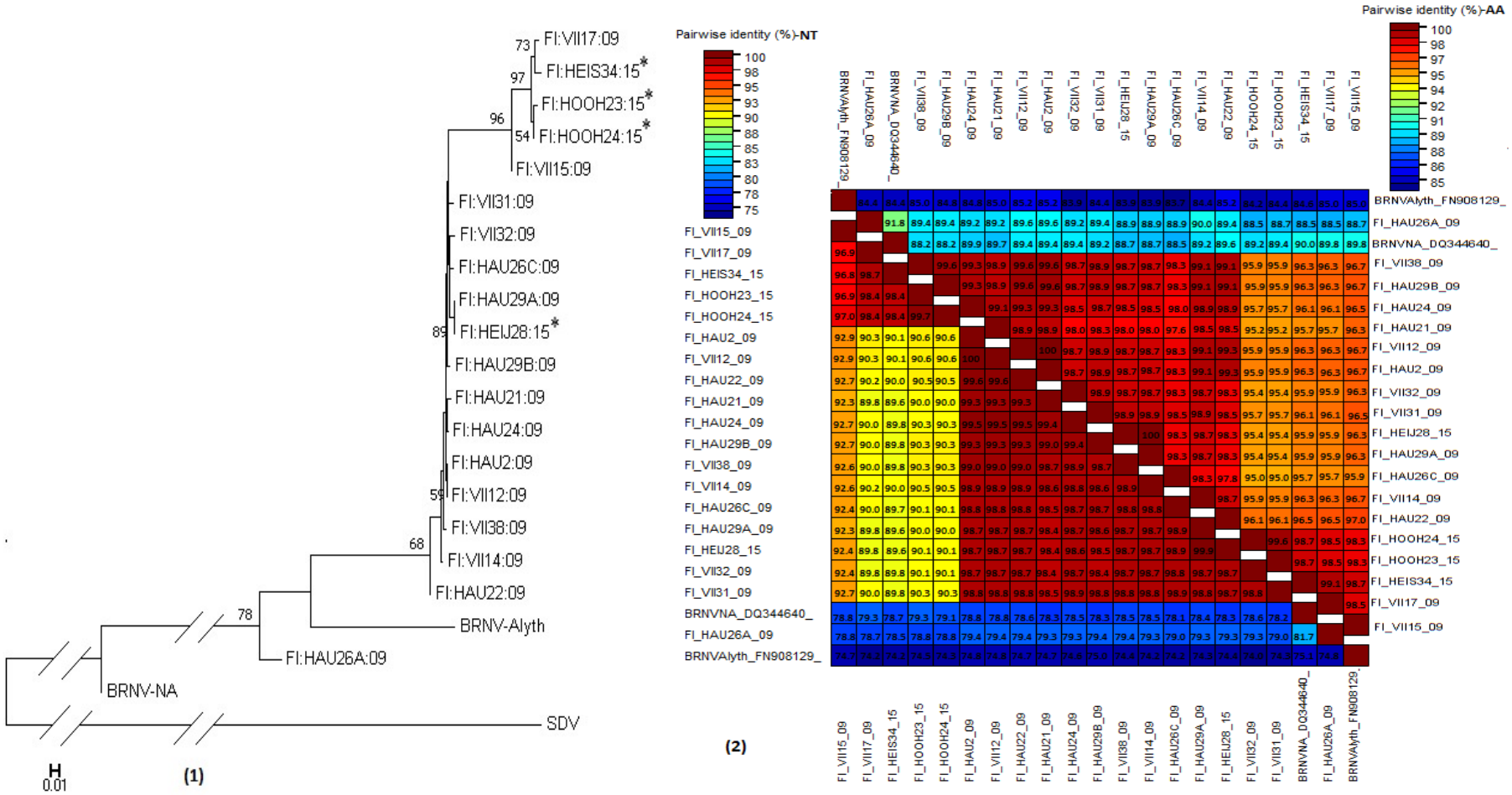


Figure 11. (1) Neighbor-joining phylogenetic analysis of small coat protein (CPs) encoding nucleotide (nt) sequences from 21 BRNV isolates rooted with SDV. Accession numbers and boot strap values >65 are shown. Kimura 2-parameter model, Gamma distributed (G). The isolates sequenced in raspberry cultivars from Finland are marked with asterisk. **(2)** Pairwise comparison of CPs-encoding nucleotide (nt) –**lower diagonal left** and CPs-encoding amino acid sequences-**upper diagonal right** of 21 BRNV isolates. The colour reference represents pairwise nucleotide and protein identities between corresponding isolates.

All isolates from Finland formed a separate branch except FI:HAU26A:09, the isolate previously sequenced from wild raspberries. This isolate formed in out-group exhibited 21% non-identity at nucleotide level (nt) and 10-12% non-identity at amino acid level (aa) to other isolates from Finland. However, it showed identity 81.7% (nt) and 91.8% (aa) with BRNV isolate from North America (DQ344640). BRNV-Alyth non-identity to all other isolates accessions used for phylogenetic analysis was found ~25% (nt) and ~15% (aa). The Phylogenetic tree was rooted with *Satsuma dwarf virus* (SDV) (Figure 11).

Restriction fragment length polymorphism (RFLP) was performed using online NEB cutter to rapidly analyse the efficiency of molecular cloning, and to probe isolate diversity of RNA viruses present in raspberry samples. Type-II *Pst*I and *Eco*RI endonuclease were selected to digest DNA sequences of isolates obtained. BRNV CPs encoding sequence digestion with *Pst*I in isolate FI:HEIJ28:15 revealed the presence of *Pst*I (CTGCA^G) site at 4069nt and 5004nt (1nt change) positions (Figure 12). This isolate was recovered from cultivar RU24 Heija and its sequence homology at nucleotide level was identified 99.9% with isolate FI:HAU29A:09, which also contains *Pst*I sites. FI:HOOH23:15, FI:HOOH24:15 and FI:HEIS34:15 are other BRNV isolates sequenced for this study, *Pst*I restriction site is absent in CPs nucleotide sequence of those isolates. *Eco*RI presence is not discussed here.

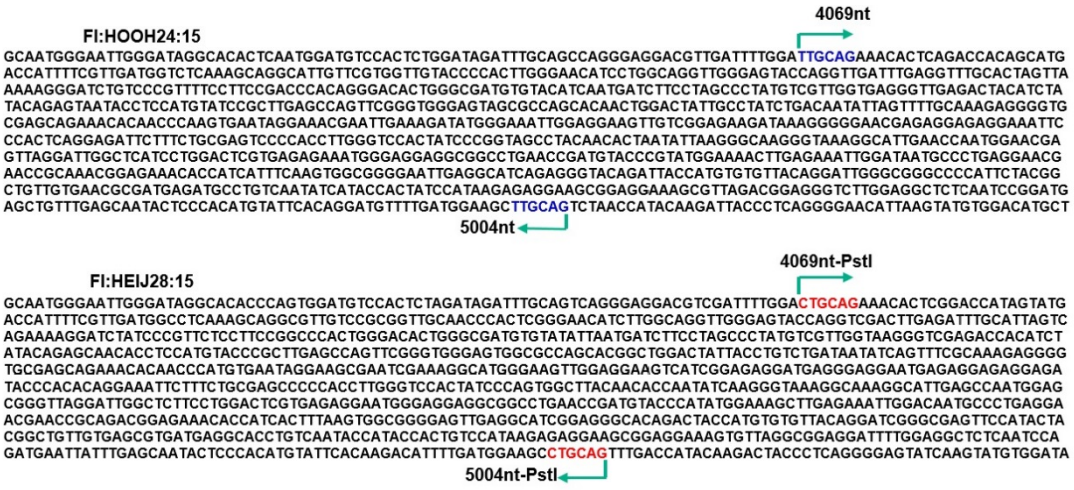


Figure 12. Sequence comparison of two BRNV isolate accessions from raspberry cultivars. FI:HEIJ28:15 possess *Pst*I restriction site at 4069nt and 5004nt whereas absent in FI:HOOH24:15. The position marking is cited of DQ344640 as reference.

5 DISCUSSION

Presence of RBDV and BRNV in HXR-2 raspberry sample pool was quickly diagnosed using VirusDetect. These two RNA viruses were detected based on homology-dependent reference guided contig assembly. VirusDetect sensitivity was reaffirmed, contigs assembled along the 3'end due to substantial read availability made BRNV detection easily possible. Raspberry host genome was not available to subtract host derived sRNA reads. Host genome subtraction would enrich virus derived reads, and can rapidly produce detection results. VirusDetect performance was previously demonstrated to detect plant, insect viruses and unidentified viruses from RNAseq data (Zheng et al. 2017). For genome mapping the initial use of VirusDetect helped to select reference genomes from list of results displayed.

Genome alignment and mapping indicated vsiRNA read size (21nt & 22nt) preponderance to cover RBDV and BRNV genomes. This clearly pointed the benchmark of PTGS that encounters RNA virus genome. DCL-4 dicing of perfect 21nt size in terms of Anti-viral silencing pathway was represented well, DCL-2 product of 22nt size presence was reaffirmed in genome maps and in regions covered by vsiRNAs. All known and novel (+) RNA viruses so far detected by sRNA deep sequencing based diagnosis showed 21nt and 22nt vsiRNAs prevalence that targeted genomes (Kreuze et al. 2009, Donaire et al. 2009, Barrero et al. 2017). Plant (+) RNA viruses replicate in cytoplasm and DNA viruses in nucleus, it was described that DNA virus transcripts encounter PTGS in cytoplasm possibly during times of viral mRNA translation (Pumplin and Voinnet 2013). RNA virus discovery had not been reported using unique 24nt vsiRNA size profile, but parareto (dsRNA-RT) viruses and DNA viruses detection included 21, 22, 24nt vsiRNAs, selective detection of dsRNA-RT virus by sorting 24nt vsiRNA size was also demonstrated (Seguin et al. 2013, Barrero et al. 2017). The read size 23nt and 24nt also covered RBDV and BRNV references but was considered redundant. Data mining studies reported vsiRNA profiles 21nt and 22nt size with perfect 2' overhang are first-rate duplexes for RNA virus discovery from bulk sRNA datasets (Niu et al. 2017). It is possible that 23nt and 24nt vsiRNA reads aligned RBDV and BRNV references were imperfect vsiRNA duplexes with extra nucleotides along overhang at 3'end. It was evident from studies that DCL-4 dices for 21nt and its surrogate DCL-2 dices for 22nt vsiRNAs and no exact 23nt vsiRNA

production from DCLs was reported (Kreuze et al. 2009, Donaire et al. 2009, Pumplin & Voinnet, 2013).

Populations of vsiRNA mapped to RBDV genomes displayed possible importance of genomic elements in 3'UTR, it was too observed as the least covered region, especially in RBDV RNA-1. Putative SL structures were previously reported at 3' UTR of RBDV RNA-1 and RNA-2. Last 18nt are identical and can base pair to form SLs (Ziegler et al. 1992). Biogenesis of vsiRNAs showed correlation to SLs, also it was related to secondary structures or fold back regions that can produce DCL targeted vsiRNAs, than other region in genomes (Donaire et al, 2009, Pumplin & Voinnet 2013, Wu et al. 2013) But, RISC associated AGOs cannot easily accesses highly structured regions, similar reported for viroids (Wu et al. 2013, Hohn 2015). In case of RBDV 3'UTR, it is possible that these structural regions were not targeted well by AGOs. In RBDV RNA-1 a putative overlapping ORF presence was previously reported. It was positioned downstream to Replicase protein and marked as protein p12, and aligns with CMV 2b gene, a widely investigated VSR (Valasevich et al. 2011). vsiRNA coverage was sporadic at this position near 3' end of RNA-1.

The hotspot indicated at the 5'end in RBDV RNA-1 and RNA-2 was mostly mapped by 21nt vsiRNAs. Towards the 3'end vsiRNA coverage was observed less, where possible sgRNA promoter and verified CP encoding sequence exists. Molecular studies had revealed the multifunctional action of RBDV CP, and its encoding sequence (sgRNA-133nts) in RNA-2 3'end. CP enhances viral replication, giving the phenomenon name 'Genome activation' (MacFarlane & McGavin 2009). Similar type of strategy occurs in viruses like *Tobacco streak virus* (TSV) and AMV, enhanced replication occurs when CP binds at base of two SL structures separated by AUGC motifs in genomic 3'UTRs. *Ilarviruses* (TSV and AMV) CP were exchangeable for genome activation but not exchangeable from or to RBDV (MacFarlane & McGavin 2009). Plant RNA virus replicates and translates basing Replicase proteins at positions in *cis*-acting elements of template, predominant regulatory elements are present near 3' end of genomes (Chujo et al. 2015, Carbonell et al. 2016).

RBDV genomic SLs can be regulators for translation and replication. Ecuadorian RBDV isolate (KJ007640) RNA-2 showed concatenation (RNA-2 + inverted sgRNA) in CP encoding region leading to 3279nt RNA-2. Novel *Idaeovirus*, *Blackcurrant leaf*

chlorosis associated virus (BLCaV) full genome was determined showing exact genome concatenation (James & Phelan 2017). Deriving global genome organization may reveal the regulatory elements present in both coding and non-coding regions of RBDV genome, also experiments with anti-viral silencing DCLs and AGOs can be examined (Wu et al. 2013). RBDV CP encoding region phylogeny verified that the virus diversity is intervened of its host plant. Red raspberry isolates were grouped together regardless of geographical origin, but in Finland its diversity was limited of geographical occurrence. Grapevine isolate of RBDV and wild raspberry were grouped outside. The low level isolate variability is mostly due to the method of red raspberry propagation (Valasevich et al. 2011).

BRNV-NA genomes mapped vsiRNA reads at 3' UTR region, and the conspicuous hot spot was marked by 22nt vsiRNA size. The preferential target of DCL-2 at genomic regions has been attributed in some viruses. *Cymbidium ringspot virus* (CymRSV) sRNA profiling accumulated 22nt vsiRNA than 21nt and the possibility of CymRSV VSR (suppressor) p19 sequestering 21nt vsiRNAs to produce 22nt vsiRNAs was discussed. Similar action had been detailed about TCV VSR p38 that interfered DCL4 dicing action to produce more 22nt vsiRNA by its surrogate DCL2 (Pumblin and Voinnet 2013). Very recent information was obtained from VSR P15 of PCV that contains peroxisome targeting signal (PTS1), which bound more 22nt vsiRNAs to peroxisome and stopped systemic spread of silencing, where 21nt vsiRNA production was also reported (Incarbone et al. 2017, Daros 2017). Studies indicate that 22nt vsiRNA has less role in viral genome target but important for systemic spread of silencing signal (Pumplin and Voinnet 2013). It may be possible to comment that the coding regions of BRNV were evaded and terminal regions were exposed. Two significant encoding regions, RdRp and CPs were calculated of less coverage by vsiRNAs and are exact upstream to 3'UTR. The sequence similarity at 3'UTR in RNA-1 and RNA-2 was reported 98% and terminal 12nt are identical, besides both 3'UTR are ~940nts in length (McGavin et al. 2010). CaMV 8s RNA is highly structured to produce massive vsiRNAs that the coding regions are not exposed to AGOs for RISC mediated coverage (Hohn 2015). If BRNV 3' UTR sequence can form adequate secondary structures, then DCL targeting of this region can be assumed.

BRNV RdRp sequence obtained from wild *Rubus canadensis* L. was reported 24% dissimilar at protein level (aa) to BRNV-NA and BRNV-Alyth, and <50% identical to

other viruses of *Secoviridae,* also widespread in wild *Rubus* (McGavin et al. 2010, Martin et al. 2013). The remarkable feature reported in BRNV genomic RNA-1 is the presence of AlkB protein domain of oxygenase superfamily. N and O atoms in RNA substrate undergoes addition of either carbon chains or single carbon (methylation). The alkylating environments inside woody plant phloem/pesticide residue can damage viral RNA, and in order survive and spread, AlkB protein repairs RNA damage by oxidative methylation, and are present in a number of plant viruses in the family *Flexiviridae*. Viral AlkB domain homologues are present in DNA genomes of bacteria, and BRNV possibly had acquired this through recombination with other virus of same co-infection as shown for *Parsnip yellow fleck virus* of *Sequiviridae*. (Halgren et al. 2007, Barr and Fearns 2010, McGavin et al. 2010) Previously, AlkB presence in viral genomes intrigued the possibility of silencing suppression, however *in vivo* studies failed to show AlkB activity on siRNA 2' O-methylation (Van den Born et al. 2008).

BRNV isolates sequenced from four cultivars exhibited diversity at nucleotide level and the presence of similar and diverse isolates in both cultivated and wild raspberries in Finland. BRNV CPs (encoding) phylogeny pointed that the isolates were grouped based on geographical origin yet highly variable isolates were present in same regions. One BRNV cultivar isolate now sequenced (FI:HEIJ28:15) phylogenetically aligned with isolates sequenced previously from wild raspberries. It is noteworthy to mention that BRNV is transmitted well among wild raspberries in Finland. North American isolate CPs sequence showed 74% identity with BRNV- Alyth. Possible recombination occurring at CPs encoding regions sequence is now speculated. Restriction endonucleases are used to distinguish strains for single and mixed infections. (Tugume et al. 2010). Restriction with *Pst*I type II restriction (online) distinguished two cultivar isolates in Finland. There may be BRNV strain types that are locally adapted to particular region as shown for SPFMV (Tugume et al. 2010). Finding recombination hotspots in CPs sequences can possibly indicate diversification of BRNV strains. Phylogenetically, BRNV RNA-2 polyprotein (P2) is more related to SMoV than SDV, analysing five Canadian isolates of SMoV revealed that CPs of SDV is not in the C-terminus of SMoV or BRNV P2 (Bhagwat et al. 2016). Some clones of isolates were identified of RNA diversity (quasispecies) as shown (Appendix. 10.2).

Barr & Fearns (2010) discussed about the possible recombination events of internal coding regions in RNA virus genomes. The usual polymerase-error at internal regions

are either cleared by exonuclease activity or RNA recombination. RNA recombination is well observed in long genome RNA viruses, especially near terminal 3' region coding and non-coding sites (Quito-Avila et al. 2012). Kutnjak et al. (2015), showed the existence of mutational cloud resulting recombination in sequences by comparing PVY derived vsiRNA and viral RNA populations. Both sample pools showed Single Nucleotide Polymorphism (SNPs) of 1.6% yet non-homologous recombination was observed in viral particle pool. Attention should be paid that neither BRNV nor its relative bipartite *Secoviroids* yield active virions during purification, yet are more recalcitrant with empty particles (Halgren et al. 2007, Kutnjak et al. 2015, Thompson et al. 2017).

Raspberry cultivar samples were listed in pairs for this study, and virus positive samples were resulted in pairs. It is possible to deduce that virus positive raspberry cultivars were susceptible to concerned virus. In Europe, RB 15 strain of RBDV had been found to overcome the resistance provided by *Bu* gene and few raspberry aphid biotypes were reported for resistance response in cultivars (Valasevich et al. 2011, Martin et al. 2013). Mixed infection of RBDV with BRNV in cultivar samples of HXR-2 was not observed yet these viruses are excellent agents in viral synergism. When both RBDV and BRNV result as virus complex in red raspberry, plant shoots are proliferated (Martin et al. 2013). BRNV titre increases when *Sobemovirus* is added or present in mixture, and ~400 fold increase in RBDV concentration had been attributed to RLMV presence for severe bushy dwarf in drupelets (McGavin & McFarlane 2009, Quito-Avila et al. 2012, Martin et al. 2013).

RT-PCR based BRNV characterization in individual HXR-2 raspberry samples had been tedious, since different primer pairs were used to recover isolates from virus positive samples. The detection of BRNV-NA and BRNV-Alyth was achieved through oligonucleotide primers, however not all isolates can possibly be detected due to its diversity (McGavin et al. 2010). CTAB-method based total RNA extraction procedure has been followed in this study, and during BRNV CPs sequence amplification non-specific primer binding to template even generated products of similar marker length. While sequencing of clones revealed that the product was amplified from host genomic region, two clonal sequence submitted for blast (nt) in *Roseaceae* database found sequence identical to Bras_G21586 (Blackberry) - sensor histidine kinase QseE (data not shown, observation). Nonetheless, the excellent matrix for characterizing virus

infecting small fruit berries, is to isolate dsRNA from sample of interest. This is straight-forward for RT-PCR amplification since replicative form of viral RNA available can give accurate first strand cDNA synthesis. Tanzetakis & Martin (2008), tirelessly formulated a protocol for efficient yielding dsRNA extraction protocol form blueberry tissues using CF-11 cellulose elution. This protocol based visualization of dsRNAs in agarose gels would expose full-length viral replicative/dsRNA genomes. It was also reported that dsRNA is thermostable and resistant to RNAse degradation, and the same method had been used for BRNV-NA and BRNV-UK full genome characterization with or without DOP-PCR (Halgren et al. 2007, Tanzetakis & Martin 2008, McGavin et al. 2010).

RBDV and BRNV control is an important consideration, since efficient control can be achieved using pathogen free virus-indexed cultivars. BRNV is one important virus of raspberry that alters the behaviour of the large European aphid. Volatiles secreted from virus infected plants attracted more aphids, and virus is transmitted in a semi-persistent manner (McMenemy et al. 2012). It was also advised not to keep *Rubus* plantings or plant products in field, at places viral infection is confirmed. This is important for BRNV since aerial vectors spread this virus, and black raspberries in North America are highly damaged by viruses vectored through raspberry aphid. A susceptible cultivar should be grown with replanting, if the need to non-compromise market value of that product. Only a sensible cultivar reveal symptoms and this is clear in case studies of *Rubus* (Lemmetty et al. 2011, Martin et al., 2013). Detection of viruses is the first step to find viruses of unknown origin to calculate preventive methods.

Automated software like VirusDetect can be used at places where rapid diagnosis are of demand. Indeed, low-titre viruses can possibly be detected using sRNA size profiles. Nevertheless, experimental validation of sRNA sequencing technology or concurrent RNA silencing based studies should be validated through relevant bio-informatics, also considering statistically sound sampling and analysis. Sequencing library should contain a control from healthy plants to validate the results. sRNA libraries thus prepared can be availed to remove the host sequences from healthy plants (Roy et al. 2013). The next question to be answered is how virus presence inference can be made considering preventive measures. Both RNAseq and sensitive vsiRNA deep sequencing detects virus yet active form the pathogen should be confirmed together with two or more biological assays. This importance is for any new pathogen

identification at post quarantine level, since NGS based detection method would also produce positive result for sample/cultivars after effective therapy (Martin et al. 2013). Quarantine laws for reducing the spread of viral pathogens can be coupled with germplasm certification programs. An example for *Rubus* certification in United States is tabulated under National Clean Plant Network (NCPN) for berries, up to G1 level (nuclear stock) testing is being made comprehensive through NGS technology (Gergerich et al. 2016).

An user-friendly detection software for reference genome assembly by aligning short reads, also connecting database significantly reduces time and increase accuracy for virus diagnosis, and the silencing product vsiRNA is sensitive than any other indicator plant and probing antibodies cultured from other organisms (Valkonen 2014). sRNA virus diagnosis can be expanded from detection to analysis of virus population structure in plant ecosystem. Aetiology behind previously described viral diseases can be rechecked with sRNA omics to reveal new agents, especially in tropical regions. The metagenomics of viruses are also evaluated under vsiRNA omics (Prabha, et al. 2013, Seguin et al. 2013). Taken together, vsiRNA based virus diagnosis leads among the high-throughput sequencing methods employed for accurate known and novel virus detection.

7 CONCLUSION

This study reported the presence of RBDV and BRNV in raspberry cultivars in Finland. VirusDetect boosted RNA virus detection in multiplexed raspberry sample pool HXR-2 through automated contigs built from vsiRNA read profile. The virus positive samples were verified by RT-PCR also confirmed RBDV and BRNV in sample pool. Genomes mapped by 21 and 22nt vsiRNAs verified anti-viral silencing abstraction and significant genomic regions were marked. Sanger sequencing CPs encoding region revealed more anticipatable diversity of BRNV virus isolates in cultivated raspberries correlative with isolates from wild raspberries in Finland. The information also represents the consequence of RNA silencing and recombination resulting quasispecies in low-titre BRNV as implications for accurate virus diagnosis, especially in berry crops like *Rubus*.

8 REFERENCES

Altschup, S.F., Gish, W., Miller, W., Myers, E.W. Lipman, D.J. 1990. Basic Local Alignment tool. Journal of Molecular Biology 215: 403-410.

Barr, J.N. & Fearns, R. 2010. How RNA viruses maintain their genome integrity. Journal of General Virology 91: 1373–1387.

Barrero, R.A., Napier, K.R., Cunnington, J., Liefting, L., Keenan, S., Frampton, R.A., Szabo, T., Bulmas, S., Hunter, A., Ward, L., Whattam, M. & Bellgard, M.I. 2017. An internet-based bioinformatics toolkit for plant biosecurity diagnosis and surveillance of viruses and viroids. BMC Bioinformatics: doi: 10.1186/s12859-016-1428-4.

Bhagwat, B., Dickison, V., Ding, X., Walker, M., Bernardy, M., Bouthillier, M., Creelman, A., DeYoung, R., Li, Y., Nie, X., Wang, A., Xiang, Y. & Sanfacon, H. 2016. Genome sequence analysis of five Canadian isolates of *Strawberry mottle virus* reveals extensive intra-species diversity and a longer RNA2 with increased coding capacity compared to a previously characterized European isolate. Archives of Virology 61: 1657–1663.

Bi, Y., Artola, K., Kurokura, T., Hytönen, T. & Valkonen, J.P.T. 2012. First report of *Raspberry leaf blotch virus* in raspberries in Finland. Plant disease 96: 1231.

Carbonell, A., Garciá, J.A., Simon-Mateo, C. & Hernandez, C. 2016. Plant virus RNA replication. In: eLS. John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0022338.

Cheong, E.J., Jeon, A.R., Kang, J. W., Mock, R., Kinard, G. & Li, R. 2014. *In vitro* elimination of *Black raspberry necrosis virus* from black raspberry (*Rubus occidentalis*). Horticultual Science 41: 95-99.

Chujo, T., Ishibashi, K., Miyashita, S. & Ishikawa M. 2015. Functions of the 5’- and 3’-untrasnlated regions of *tobamovirus* RNA. Virus Research 206: 82-89.

Corpet, F. 1998. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Research 16:10881-10890.

Daros, J-A. 2017. Combating RNA silencing. Nature Plants 3: 17098.

Donaire, L., Wang, L., Gonzalez-Ibeas, D., Mayer, K.F., Aranda, M.A. & Llave, C. 2009. Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. Virology 392: 203–214.

Flobinus, A., Hleibieh, K., Klein, E., Ratti, C., Bouzoubaa, S. & Gilmer, D. 2016. A viral noncoding RNA complements a weak end viral RNA silencing suppressor and promotes efficient systemic infection. Viruses 8: 272.

Ghoshal, B. & Sanfacon, H. 2015. Symptom recovery in virus-infected plants: Revisiting the role of RNA silencing mechanisms. Virology 479-489: 167-179.

Gergerich, R.C., Tzanetakis, I.E. & Martin. R.R. 2016. Towards a national certification scheme for Rubus in the United States Acta Horticulturae: DOI10.17660/ActaHortic.2016.1133.74.

Halgren, A., Tzanetakis, I.E. & Martin, R.R. 2007. Identification, characterization and detection of *Black raspberry necrosis virus*. Phytopathology 97:44-50.

Harrison, B.D., Murant, A.F, 1996. The Plant viruses: polyhedral virions and bipartite RNA genomes, Springer Press.

Ho, T. & Tzanetakis, I.E. 2014. Development of a virus detection and discovery pipeline using next generation sequencing. Virology 471-473: 54-60.

Hohn, T. 2015. RNA based viral silencing suppression in plant pararetroviruses. Frontiers in Plant Science 6: 398. doi: 10.3389/fpls.2015.00398.

Incarbone, M., Zimmermann, A., Hammann, P., Erhardt, M., Michel, F. & Dunoyer, P. 2017. Neutralization of mobile antiviral small RNA through peroxisomal support. Nature Plants 3: 17094. doi:10.1038/nplants.2017.94.

Isogai, M., Yoshida, T., Nakanowatari, C. & Yoshikawa, N. 2014. Penetration of pollen tubes with accumulated *Raspberry bushy dwarf virus* into stigmas is involved in initial of maternal tissue and horizontal transmission. Virology 452–453: 247–253.

Ivanov, K.I. & Mäkinen, K. 2012. Coat proteins, host factors and plant viral replication. Current Opinion in Virology 2:712–718.

James, D. & Phelan, J. 2017. Complete genome sequence and analysis of *Blackcurrant leaf choloris associated virus*, a new member of the genus *Idaeovirus*. Archives of Virology 162: 1705-1709.

Kallio, M.A., Tuimala, J.T., Hupponen, T., Petri Klemelä, P., Gentile, M., Scheinin, I., Koski, M., Käki, J. & Korpelainen, E.I. 2011. Chipster: use-friendly analysis software for microarray and other high-throughput data. BMC Genomics 12: 50.

Kokko, H.I., Kivineva, M., Kärenlampi, S.O, 1996. Single-step immunocapture-PCR in the detection of *Raspberry bushy dwarf virus*. Biotechniques 20: 842-846.

Kreuze, J., Perez, A., Untiveros, M., Quispe, D., Feuntes, S., Barker, I. & Simon, R. 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. Virology 388: 1-7.

Kumar, S., Stecher, G. & Tamura, K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33: 1870–1874. doi:10.1093/molbev/msw054.

Kutnjak, D., Rupar, M., Gutierrez- Aguirre, I., Curk, T., Kreuze, F.J. & Ravinkar, R. 2015. Deep sequencing of virus derived small interfering RNAs and RNAs from viral particles shows highly similar mutational landscape of a plant virus population. Journal of Virology 89: 4760-4769.

Langmead, B. & Salzberg, S. 2012. Fast gapped-read alignment with Bowtie2. Nature Methods 9: 357–359.

Lemmetty, A., Laamenen, J., Soukainen, M. & Tegel, J. 2011. Emerging virus and viroid pathogen species identified for the first time in horticultural plants in Finland. Agricultural and Food Science 20: 29-41.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. & 1000 Genome Project Data Processing Subgroup. 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25: 2078-9.

Liu, J., Zhang, X., Yang, Y., Hong, N., Wang, G., Wang, A. & Wang, L. 2016. Characterization of virus-derived small interfering RNAs in *Apple stem grooving*

virus-infected in vitro-cultured *Pyrus pyrifolia* shoot tips in response to high temperature treatment. Virology Journal 13:1-11.

MacFarlane, S.A. & McGavin, W.J. 2009. Genome activation by *Raspberry bushy dwarf virus* coat protein. Journal of General Virology 90:747-753.

Martin, R.R., MacFarlane, S.A., Sabanadzovic, S., Quito, D., Poudel, B. & Tzanetakis, I.E. 2013. Viruses and virus diseases of *Rubus*. Plant Disease 97: 168-182.

Martin, R.R., Constable, F., Tzanetakis, I.E. 2016. Quarantine regulations and the impact of modern detection methods. Annual reviews of Phytopathology 54: 189-205.

Mathews, H., Keller, K.E. & Martin, R.R. 2004. Development of resistance to *Raspberry bushy dwarf virus* in 'Meeker' red raspberry. Acta Horticulturae 656: 159-163.

McGavin, W.J. & MacFarlane, S.A. 2009. Rubus chlorotic mottle virus, a new *Sobemovirus* infecting raspberry and bramble. Virus Research 139: 10-13.

McGavin, W.J., McMenemy, L.S. & MacFarlane, S.A. 2010. The complete sequence of a UK strain of *Black raspberry necrosis virus*. Journal of General Virology 93: 430-437. Archives of Virology 155: 1897- 1899.

McMenemy, L.S., Hartley, S.E., MacFarlane, S.A., Karley, A.J., Shepherd, T. & Johnson, S.N. 2012. Raspberry viruses manipulate the behaviour of their insect vectors. Entomologia Experimentalis et Applicata 144:56-68.

Muhire, B. M, Varsani, A. & Martin, D. P. 2014. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. PLoS ONE 9: e108277.

Niu, X., Sun, Y., Chen, Z., Li, R., Padmanaban, C., Ruan, J., Kreuze, J.F., Ling, K., Fei, Z. & Gao, S. 2017. Using small RNA-seq data to detect siRNA duplexes induced by plant viruses. Genes 8:163.

Nordenstedt, N., Marcenaro, D., Chilagane, D., Mwaipopo, B., Rajamäki, M.L., Nchimbi-Msolla, S., Njau, P.J.R., Mbanzibwa, D. & Valkonen, J.P.T. 2017. Pathogenic seed-borne viruses are rare but *Phaseolus vulgaris* Endornaviruses are

common in bean varieties grown in Nicaragua and Tanzania. PLoS ONE 12 (5): e0178242. https://doi.org/10.1371/journal.pone.0178242.

Parikka, P., Tuovinen, T., Lemmetty, A. 2016. Challenges for plant protection of berry crops in northern Europe. Acta 660: ISHS. DOI 10.17660.

Prabha, K., Baranwal, V.K. & Jain, R.K. 2013. Applications of next generation high throughput sequencing technologies in characterization, discovery and molecular interaction of plant viruses. Indian Journal of Virology 24: 157-165.

Pumplin, N. & Voinnet, O. 2013. RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-counter-defence. Nature Reviews Microbiology 11: 745-760.

Quinlan, A.R., & Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842.

Quito-Avila, D.F., Martin, R.R. 2012. Real-time RT-PCR for detection of *Raspberry bushy dwarf virus*, *Raspberry leaf mottle virus* and characterizing synergistic interactions in mixed infections. Journal of Virological Methods 179: 38-44.

Quito-Avila, D.F., Lightle, D., Martin, R.R. 2014a. Effect of *Raspberry bushy dwarf virus*, *Raspberry leaf mottle virus* and *Raspberry latent virus* on plant growth and fruit crumbliness in ‘‘Meeker’’ red raspberry. Plant Disease 98: 176-183.

Quito-Avila, D.F., Ibarra, M.A. Alvarez., Peralta, E.L. & Martin, R.R. 2014b. A *Raspberry bushy dwarf virus* isolate from Ecuadorean *Rubus glaucus* contains an additional RNA that is a rearrangement of RNA-2. Archives of Virology 159: 2519-2521.

Roy, A., Shao, J., Hartung, J.S., Schneider, W. & Brlansky, R.H. 2013. A case study on discovery of novel *Citrus leprosis virus* cytoplasmic type 2 utilizing small RNA libraries by next generation sequencing and bioinformatic analyses. Journal of Data Mining in Genomics & Proteomics 4: 129. doi:10.4172/2153-0602.1000129.

Seguin, J., Rajeswaran, R., Malpica-Lopez, N., Martin, R.R., Kasschau, K., Dolja, V.V., Otten, P., Farinelli, L. & Pooggin, M.M. 2014. *De novo* reconstruction of

consensus master genomes of plant RNA and DNA viruses from siRNAs. PLoS ONE 9(2): e88513. doi:10.1371/journal.pone.0088513.

Schneider, L.W. & Roossinck, J.M. 2001. Genetic diversity in RNA quasispecies is controlled by host-virus interactions. Journal of Virology 75: 6566–6571.

Thompson, J.R, Dasgupta, I., Fuch, M., Iwanami, I., Karasev, A.V., Petrzik, K., Sanfacon, H., Tzanetakis, I. E., Van der Vlugt, R, Wetzel, T., Yoshikawa, N. & ICTV Report Consortium. 2017. ICTV virus taxonomy profile: *Secoviridae.* Journal of General Virology. doi: 10.1099/jgv.0.000779.

Taito terminal. 2013. CSC – IT Center for Science, Finland, for computational resources https://research.csc.fi/taito-user-guide.

Tian, Y-P., Hepojoki, J., Ranki, H., Lankinen, H., Valkonen, J.P.T. 2014. Analysis of *Potato virus Y* coat protein epitopes recognized by three commercial monoclonal antibodies. PLoS ONE 9: e115766.

Tzanetakis, I.E. & Martin, R.R. 2008. A new method for extraction of double-stranded RNA from plants. Journal of Virological methods 149: 167-170.

Tzanetakis, I. E. 2012. The Gordian knot of small fruit virology: Emerging diseases and their control. APS Features. doi: 10.1094/APSFeature-2012-01.

Tugume, A.K., Cuellar, W. I., Mukasa, S.B. & Valkonen, J.P.T. 2010. Molecular genetic analysis of virus isolates from wild and cultivated plants demonstrates that East Africa is a hotspot for the evolution and diversification of *Sweet potato feathery mottle virus*. Molecular Ecology 19: 3139-3156.

Van den Born, E., Omelchenko, M.V., Bekkelund, A., Leihne, V., Koonin, E.V., Dolja, V.V. & Falnes, P.O. 2008. Viral AlkB proteins repair RNA damage by oxidative demethylation. Nucleic Acid Research 36: 5451-5461.

Valasevich, N., Kukharchyk, N. & Kvarnheden, A. 2011. Molecular characterization of *Raspberry bushy dwarf virus* isolates from Sweden and Belarus. Archives of Virology 156: 369-374.

Valkonen J.P.T. 2007. Viruses: economical losses and biotechnical potential. In: Potato Biology and Biotechnology: Advances and Perspectives. Vreugdenhill, D. (ed.) pp. 619- 641. Elsevier Science, Amsterdam, the Netherlands.

Valkonen, J.P.T. 2014. The challenges of controlling plant viruses. Annals of Applied Biology 164: 313-317.

Vincze, T., Posfai, J. & Roberts, R.J. 2003. NEBcutter: a program to cleave DNA with restriction enzymes. Nucleic Acids Research 31: 3688-3691.

Wang, Q., Cuellar, W.J., Rajamäki, M.L., Hirata, Y. & Valkonen, J.P.T. 2008. Combined thermotherapy and cryotherapy for efficient virus eradication: relation of virus distribution, subcellular changes, cell survival and viral RNA degradation in shoot tips. Molecular Plant Pathology 9: 237-250.

Weinheimer, I., Jiu, Y., Rajamäki, M-L., Matilainen, O., Kallijärvi, J., Cuellar, W.J., Lu, R., Saarma, M., Holmberg, C.I., Jäntti, J. & Valkonen, J.P.T. 2015. Suppression of RNAi by dsRNA-Degrading RNaseIII enzymes of viruses in animals and plants. PLoS Pathogens 11(3): e1004711.

Wu, B., Grigull, J., Ore, M.O., Morin, S. & Andrew White, K. 2013. Global Organization of a positive-strand RNA virus genome. PLOS pathogens 9: e1003363.

Zheng, Y., Gao, S., Padmanabhan, C., Li, R., Galvez, M., Gutierrez, D., Fuentes, S., Ling, K-S., Kreuze, J & Fei, Z. 2017. VirusDetect: An automated pipeline for efficient virus discovery using deep sequencing of small RNAs. Virology 500: 130-138.

Ziegler, A., Natsuaki, T., Mayo, M.A., Jolly, C.A. & Murant, A.F. 1992. The nucleotide sequence of RNA-1 of raspberry bushy dwarf virus. Journal of General Virology 73: 3213-3218.

9 ACKNOWLEDGEMENT

I am privileged for receiving guidance under Prof. Jari Valkonen and KPAT team during whole degree programme. Professor consistently welcomed study ideas for improving, flexibly guided thesis work and patiently answered all queries. The mediation of Dr. Minna Rajamäki had been essential for practical virology courses, including thesis work. Her advice is quick and encouraged reading articles, this helped writing various scientific reports. Technical assistance was complete, Katrin Artola supported the courage to work independently, pointing things of caution in laboratory. Thanks to professor, Heikki Hokkanen of Agricultural zoology and Asko Hannukkala of Plant pathology, for I gained significant skill oriented knowledge from those courses. Thomas Jenner Augustine gave advice concerning sRNA read extraction, alignment to reference genome. Erasmus Mundus-BRAVE scholarship provided for Master-degree mobility is acknowledged well, kind regards to all project members and international student officers at University of Helsinki. Friends during laboratory days to chat- Angana, Archana, Elina, Nandita, Shiva & Ville, but Pinky & Rohan engaged by discussion of studies. I found Aadhavan and Johannes at Latokartonon student village, and we served one another occasionally. Viikki campus arboretum, natural reserve provided refuge even during snowy winter, also thrilled me in time of birds, berries and bumblebees.

Sabu Kiriyapillai is always there for support with advice, who introduced me to the normal life in Finland. Thanks for good-natured families- Mäki, Lakanen & Piipponen gave enough memories, Vesa and Anna Mäki entertained a lot. Lasting support I receive from my family, thanks are due to parents and brothers for enduring my very long absence, yet believing strongly. We missed Aunt and James, may their memory be a blessing in lives of us. Greetings to all my teachers, friends and families in Vellore and Thiruvananthapuram. Thanks to a mysterious personality with sagacious attitude in Bioethics course, also leading me to the secret of brain and immunity.

10 APPENDIX

10.1 Sequence information

Fasta format nucleotide sequence information of BRNV CPs isolates recovered from 4 raspberry cultivars. The Accession numbers issued are used as marks.

>FI:HOOH23:15
GCAATGGGAATTGGGATAGGCACACTCAATGGATGTCCACTCTGGATAGATTTGCAGCTAGGGAG
GACGTTGATTTTGGATTGCAGAAACACTCAGACCACAGCATGACCATTTTCGTTGATGGTCTCAAAG
CAGGCATTGTTCGTGGTTGTACCCCACTTGGGAACATCCTGGCAGGTTGGGAGTACCAGGTTGATT
TGAGGTTTGCACTAGTTAAAAAGGGATCTGTCCCGTTCTCCTTCCGACCCACAGGGACACTGGGCG
ATGTGTACATCAATGATCTTCCTAGCCCTATGTCGTTGGTGAGGGTTGAGACTACATCTATACAGAG
TAATACCTCCATGTATCCGCTTGAGCCAGTTCGGGTGGGAGTAGCGCCAGCACAACTGGACTATTG
CCTATCTGACAATATTAGTTTTGCAAAGAGGGGTGCGAGCAGAAACACAACCCAAGTGAATAGGAA
ACGAATTGAAAGATATGGGAAATTGGAGGAAGTTGTCGGAGAAGATGAAGGGGGAACGAGAGGA
GAGGAAATTCCCACTCAGGAGATTCTTTCTGCGAGTCCCCACCTTGGGTCCACTATCCCGGTAGCCT
ACAACACTAATATTAAGGGCAAGGGTAAAGGCATTGAACCAATGGAACGAGTTAGGATTGGCTCAT
CCTGGATTCGTGAGAGAAATGGGAGGAGGCGGCCTGAACCGATGTACCCGTATGGAAAACTTGAG
AAATTGGATAATGCCCTGAGGAACGAACCGCAAACGGAGAAACACCATCATTTCAAGTGGCGGGG
AATTGAGGCATCAGAGGGTACAGATTACCATGTGTGTTACAGGATTGGGCGGGCCCCATTCTACGG
CTGTTGTGAACGCGATGAGATGCCTGTCAATATCATACCACTATCCATAAGAGAGGAAGCGGAGGA
AAGCGTTAGACGGAGGGTCTTGGAGGCTCTCAATCCGGATGAGCTGTTTGAGCAATACTCCCACAT
GTATTCACAGGATGTTTTGATGGAAGCTTGCAGTCTAACCATACAAGATTACCCTCAGGGGAACATT
AAGTATGTGGACATGCTGGATGGAGTGCTCAAGGAATTCGGAGTGAACATCCGGGATTACAAGTA
CAGGGTGAACTGCGGTTGCTGTGGAAAGAAAGAGATGGTCCTCCAATGCAATAAAGGATGGCGCC
TTAATGCACTTAAACGTGGTGGTTTCTCCGTGCGGGAAGGAAGAAGGACAGGCTTAGATTACCTCG
GGTCACCATCATGGGCTGGCGGAGTTATTGGTATCACTAGCGATCTGACCCCAGTCGTCACACGGC
AGGCCTTCGGAGAGCTAGTTGATCAGGTTGCCGAGAGCATGTTTGCTGAGAGCAAAAATAATGTAC
CTGGGTATCTCAAGGAGCTTGAGAAGTTCCACCC
>FI:HOOH24:15
GCAATGGGAATTGGGATAGGCACACTCAATGGATGTCCACTCTGGATAGATTTGCAGCCAGGGAG
GACGTTGATTTTGGATTGCAGAAACACTCAGACCACAGCATGACCATTTTCGTTGATGGTCTCAAAG
CAGGCATTGTTCGTGGTTGTACCCCACTTGGGAACATCCTGGCAGGTTGGGAGTACCAGGTTGATT
TGAGGTTTGCACTAGTTAAAAAGGGATCTGTCCCGTTTTCCTTCCGACCCACAGGGACACTGGGCG
ATGTGTACATCAATGATCTTCCTAGCCCTATGTCGTTGGTGAGGGTTGAGACTACATCTATACAGAG
TAATACCTCCATGTATCCGCTTGAGCCAGTTCGGGTGGGAGTAGCGCCAGCACAACTGGACTATTG
CCTATCTGACAATATTAGTTTTGCAAAGAGGGGTGCGAGCAGAAACACAACCCAAGTGAATAGGAA
ACGAATTGAAAGATATGGGAAATTGGAGGAAGTTGTCGGAGAAGATAAAGGGGGAACGAGAGGA
GAGGAAATTCCCACTCAGGAGATTCTTTCTGCGAGTCCCCACCTTGGGTCCACTATCCCGGTAGCCT
ACAACACTAATATTAAGGGCAAGGGTAAAGGCATTGAACCAATGGAACGAGTTAGGATTGGCTCAT
CCTGGACTCGTGAGAGAAATGGGAGGAGGCGGCCTGAACCGATGTACCCGTATGGAAAACTTGAG
AAATTGGATAATGCCCTGAGGAACGAACCGCAAACGGAGAAACACCATCATTTCAAGTGGCGGGG
AATTGAGGCATCAGAGGGTACAGATTACCATGTGTGTTACAGGATTGGGCGGGCCCCATTCTACGG
CTGTTGTGAACGCGATGAGATGCCTGTCAATATCATACCACTATCCATAAGAGAGGAAGCGGAGGA

AAGCGTTAGACGGAGGGTCTTGGAGGCTCTCAATCCGGATGAGCTGTTTGAGCAATACTCCCACAT
GTATTCACAGGATGTTTTGATGGAAGCTTGCAGTCTAACCATACAAGATTACCCTCAGGGGAACATT
AAGTATGTGGACATGCTGGATGGAGTGCTCAAGGAATTCGGAGTGAACATCCGGGATTACAAGTA
CAGGGTGAACTGCGGTTGCTGTGGAAAGAAAGAGATGGTCCTCCAATGCAATAAAGGATGGCGCC
TTAATGCACTTAAACGTGGTGGTTTCTCCGTGCGGGAAGGAAGAAGGACAGGCTTAGATTACCTCG
GGTCACCATCATGGGCTGGCGGAGTTATTGGTATCACTAGCGATCTGACCCCAGTCGTCACACGGC
AGGCCTTCGGAGAGCTAGTTGATCAGGTTGCCGAGAGCATGTTTGCTGAGAGCAAAAATAATGTAC
CTGGGTATCTCAAGGAGCTTGAGAAGTTCCACCC

> FI:HEIJ28:15

GCAATGGGAATTGGGATAGGCACACCCAGTGGATGTCCACTCTAGATAGATTTGCAGTCAGGGAG
GACGTCGATTTTGGACTGCAGAAACACTCGGACCATAGTATGACCATTTTCGTTGATGGCCTCAAAG
CAGGCGTTGTCCGCGGTTGCAACCCACTCGGGAACATCTTGGCAGGTTGGGAGTACCAGGTCGACT
TGAGATTTGCATTAGTCAGAAAAGGATCTATCCCGTTCTCCTTCCGGCCCACTGGGACACTGGGCGA
TGTGTATATTAATGATCTTCCTAGCCCTATGTCGTTGGTAAGGGTCGAGACCACATCTATACAGAGC
AACACCTCCATGTACCCGCTTGAGCCAGTTCGGGTGGGAGTGGCGCCAGCACGGCTGGACTATTAC
CTGTCTGATAATATCAGTTTCGCAAAGAGGGGTGCGAGCAGAAACACAACCCATGTGAATAGGAA
GCGAATCGAAAGGCATGGGAAGTTGGAGGAAGTCATCGGAGAGGATGAGGGAGGAATGAGAGG
AGAGGAGATACCCACACAGGAAATTCTTTCTGCGAGCCCCCACCTTGGGTCCACTATCCCAGTGGCT
TACAACACCAATATCAAGGGTAAAGGCAAAGGCATTGAGCCAATGGAGCGGGTTAGGATTGGCTC
TTCCTGGACTCGTGAGAGGAATGGGAGGAGGCGGCCTGAACCGATGTACCCATATGGAAAGCTTG
AGAAATTGGACAATGCCCTGAGGAACGAACCGCAGACGGAGAAACACCATCACTTTAAGTGGCGG
GGAGTTGAGGCATCGGAGGGCACAGACTACCATGTGTGTTACAGGATCGGGCGAGTTCCATACTA
CGGCTGTTGTGAGCGTGATGAGGCACCTGTCAATACCATACCACTGTCCATAAGAGAGGAAGCGG
AGGAAAGTGTTAGGCGGAGGATTTTGGAGGCTCTCAATCCAGATGAATTATTTGAGCAATACTCCC
ACATGTATTCACAAGACATTTTGATGGAAGCCTGCAGTTTGACCATACAAGACTACCCTCAGGGGA
GTATCAAGTATGTGGATATGCTGGATGGTGTGCTTAAAGAATTCGGAGTGAACATCCGGGATTACA
AGTACAGGGTGAACTGTGGTTGTTGTGGGAAGAAAGAGATGGTCCTCCAATGCAATAAAGGATGG
CGCCTTAATGCACTCAAGCGTGGTGGTTTCTCTGTGCGGGAAGGAAGGAGGACAGGCTTGGATTAC
CTCGGGTCACCATCATGGGCTAGTGGAGTTATTGGTATCACTAGCGATCTGACTCCAGTTGTTACAC
GACAGGCCTTTGGAGAGCTAGTCGACCAGGTTGCTGAGAGCATGTTCGCTGAAAGCAAAAACAAT
GTACCTGGGTATCTCAAGGAGCTTGAGAAGTTCCACCC

>FI:HEIS34:15

GCAATGGGAATTGGGATAGGCACACTCAATGGATGTCCACTCTGGATAGATTTGCAGCTAGGGAG
GACGTTGACTTTGGATTGCAGAAACACTCAGACCACAGCATGACCATTTTCGTTGATGGTATCAAAG
CAGGCATTGTTCGTGGTTGTACCCCACTTGGGAACATCCTGGCAGGTTGGGAGTACCAGGTTGATT
TGAGGTTTGCACTAGTTAGAAAGGGATCTATCCCGTTTTCCTTCCGACCCACAGGGACACTGGGCG
ATGTGTACATCAATGATCTTCCTAGCCCTATGTCGTTGGTGAGGGTTGAGACTACATCCATACAGAG
TAATACCTCCATGTATCCGCTTGAGTCAGTTCGGGTGGGAGTAGCGCCAGCACAACTGGACTATTAC
CTATCTGACAATATTAGTTTTGCAAAGAGGGGCGCGAGCAGAAACACAACCCAAGTGAATAGGAA
ACGAATTGAAAGATATGGGAAATTGGAGGAAGTTGTCGGAGAAGATGAAGGGGGAACGAGAGGA
GAGGAAATTCCCACTCAGGAGATTCTTTCTGCGAGTCCCCACCTTGGGTCCACTATCCCGGTAGCCT
ACAATACTAATATTAAGGGCAAGGGTAAAGGCATTGAACCAATGGAACGAGTTAGGATTGGCTCGT
CCTGGACTCGTGAGAGAAACGGGAGGAGGCGGCCTGAACCGATGTACCCGTATGGAAAACTTGAG
AAATTGGATAATGCCCTGAGGAACGAACCGCAAACGGAGAAACACCATCATTTCAAGTGGCGGGG

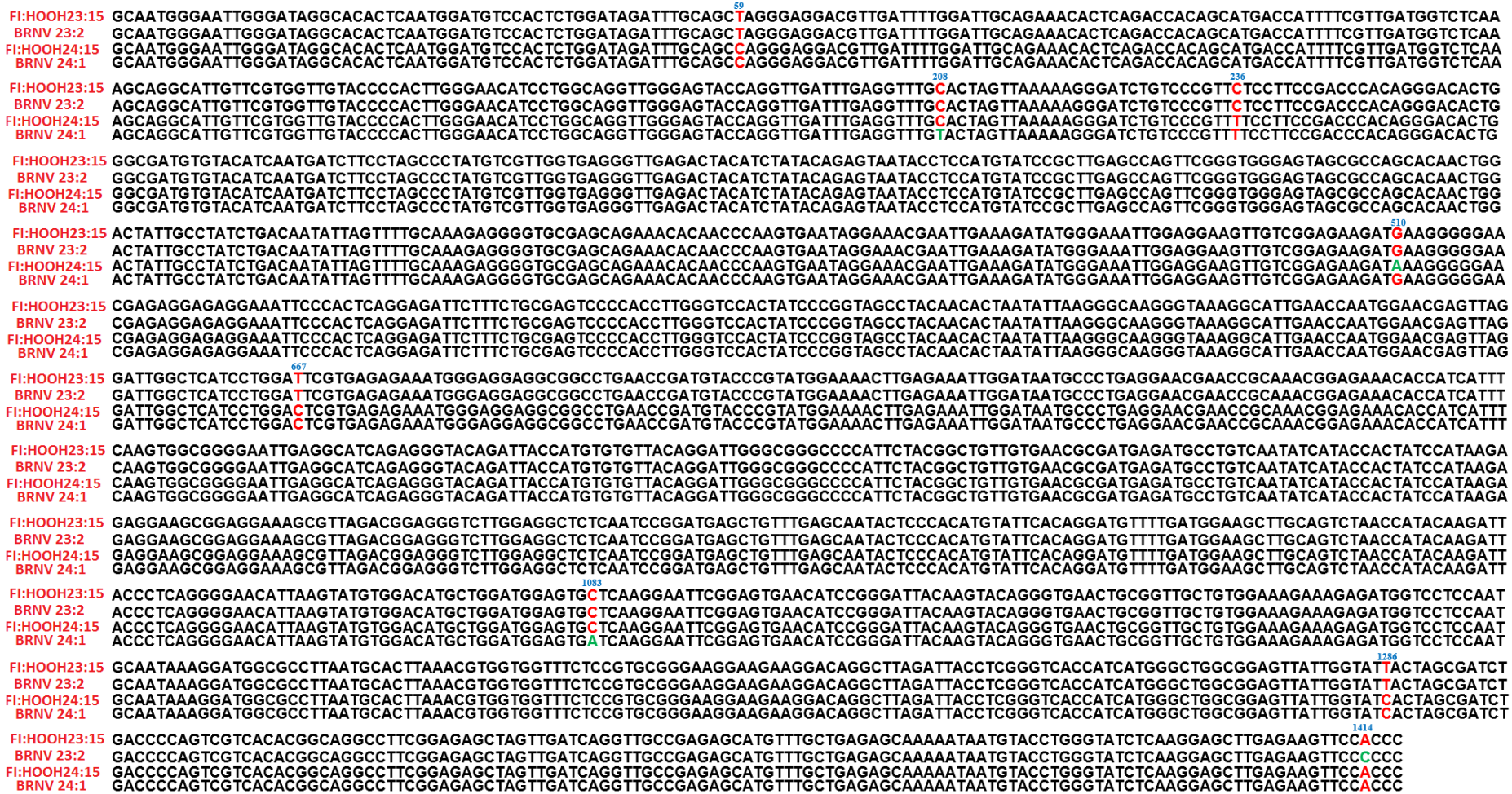
AATTGAGGCATCAGAGGGTACAGATTATCATGTGTGTTACAGGATTGGGCGGGCCCCATTCTACGG
CTGTTGTGAACGCGATGAGATGCCTGTCAATATCATACCACTATCCATAAGAGAGGAAGCGGAGGA
AAGCGTTAGGCGGAGGGTCTTGGAGGCTCTCAACCCGGATGAGCTGTTTGAGCAATACTCCCACAT
GTATTCACAGGATGTTTTGATGGAAGCTTGCAGTCTAACCATACAAGATTACCCTCAGGGGAACATT
AAGTATGTGGACATGCTGGATGGAGTGCTCAAGGAATTCGGGGTGAATATCCGGGATTACAAGTA
CAGGGTGAACTGTGGTTGCTGTGGGAAGAAAGAGATGGTCCTCCAATGCAATAAAGGATGGCGTC
TTAATGCACTTAAACGTGGTGGTTTCTCCGTGCGGGAAGGAAGAAGGACAGGCTTAGACTACCTCG
GGTCACCATCATGGGCTGGCGGAGTTATTGGTATCACTAGCGATCTGACCCCAGTCGTCACACGGC
AGGCCTTCGGAGAGCTAGTTGATCAGGTTGCCGAGAGCATGTTTGCTGAGAGCAAAAATAATGTAC
CTGGGTATCTCAAGGAGCTTGAGAAGTTCCACCC

Fasta format information of RBDV CP encoding nucleotide sequence recovered from raspberry cultivar Ville (sample no.38) of pool HXR-2.

> RBDV 38
GCTGTTCCACCAATCGTTAAGGCTCAATACGAGCTTTATAATCGTAAGTTGAACAGAGCCATCAAGG
TTTCCGGCAATCAGAAGAAGCTGGACGCTTCTTTTGTCGGGTTCAGTGAGAGCTCTAACCCAGAAA
CTGGGAAACCCCATGCGGACATGTCTATGTCTGCTAAGGTTAAGCGCGTTAATACGTGGCTTAAGA
ACTTTGATCGCGAATATTGGGATAACCAATTCGCATCAAAACCCGTTCCTAGACCTGCGAAACAGGT
CTTGAAGGGTTCATCCTCCAAATCTCAACAACGAGATGAAGGAGAGGTGGTCTTTACCCGAAAAGA
CTCCCAGAAATCCGTTAGGACTGTGTCCTATTGGGTTTGTACTCCTGAGAAGTCAATGAAACCTCTC
AAATACAAGGAGGACGAAAACGTCGTTGAAGTTACCTTCAATGACCTCACAGCTCAAAAAGCTGGT
GATAAATTGGTTTCCATCCTGTTGGAAATCAATGTGGTGGGCGGTGCCGTCGACGACAAGGGTCGA
GTGGCTGTCCTGGAAAAGGATGCTGCCGTGACGGTTGATTACCTTCTCGGAAGCCCGTATGAAGCC
ATAAATCTTGTTTCGGGTTTAAACAAGATAAATTTTAGGTCCATGACGGATGTGGTAGATTCCATAC
CATCGCTCTTAAATGAGCGTAAGGTGTGTGTCTTCCAGAATGACGACAGTTCGTCGTTTTACATTCG
GAAATGGGCTAACTTCCTTCAGGAAGTTTCGGCAGTTTTACCGGTAGGAACCGGTAAATCTTCCACA
ATAGTTCTAACTTAGTTCGGACTATTTGCCTGCTAGTGGGCATATCTCACAATTTGTGAGATAAACG
GCATAGC

10. 2 BRNV Quasispecies

BRNV CPs (encoding region) cloning and sequencing (1417nt) exhibited nucleotide non-identity or nucleotide mismatch. The clones of isolate were generated from *E.coli* colonies of same PCR product. The example given below shows the isolate FI:HOOH23:15 and other clone of same sample, BRNV 23:2. The nucleotide positions changed are marked (considering 1417nt as length). The isolate/clone difference of nucleotide is marked in red colour font, ‘Quasispecies’ or viral RNA diversity is marked in green colour font. For accuracy, another isolate FI:HOOH24:15 and its related clone of same sample, BRNV 24:1 is shown, change occurs at the same nucleotide position in all isolates/clones.


59
FI:HOOH23:15 GCAATGGGAATTGGGATAGGCACACTCAATGGATGTCCACTCTGGATAGATTTGCAGCTAGGGAGGACGTTGATTTTGGATTGCAGAAACACTCAGACCACAGCATGACCATTTTCGTTGATGGTCTCAA
BRNV 23:2 GCAATGGGAATTGGGATAGGCACACTCAATGGATGTCCACTCTGGATAGATTTGCAGCTAGGGAGGACGTTGATTTTGGATTGCAGAAACACTCAGACCACAGCATGACCATTTTCGTTGATGGTCTCAA
FI:HOOH24:15 GCAATGGGAATTGGGATAGGCACACTCAATGGATGTCCACTCTGGATAGATTTGCAGCCAGGGAGGACGTTGATTTTGGATTGCAGAAACACTCAGACCACAGCATGACCATTTTCGTTGATGGTCTCAA
BRNV 24:1 GCAATGGGAATTGGGATAGGCACACTCAATGGATGTCCACTCTGGATAGATTTGCAGCCAGGGAGGACGTTGATTTTGGATTGCAGAAACACTCAGACCACAGCATGACCATTTTCGTTGATGGTCTCAA
208
236
FI:HOOH23:15 AGCAGGCATTGTTCGTGGTTGTACCCCACTTGGGAACATCCTGGCAGGTTGGGAGTACCAGGTTGATTTGAGGTTTGCACTAGTTAAAAAGGGATCTGTCCCGTTCTCCTTCCGACCCACAGGGACACTG
BRNV 23:2 AGCAGGCATTGTTCGTGGTTGTACCCCACTTGGGAACATCCTGGCAGGTTGGGAGTACCAGGTTGATTTGAGGTTTGCACTAGTTAAAAAGGGATCTGTCCCGTTCTCCTTCCGACCCACAGGGACACTG
FI:HOOH24:15 AGCAGGCATTGTTCGTGGTTGTACCCCACTTGGGAACATCCTGGCAGGTTGGGAGTACCAGGTTGATTTGAGGTTTGCACTAGTTAAAAAGGGATCTGTCCCGTTTTCCTTCCGACCCACAGGGACACTG
BRNV 24:1 AGCAGGCATTGTTCGTGGTTGTACCCCACTTGGGAACATCCTGGCAGGTTGGGAGTACCAGGTTGATTTGAGGTTTGTACTAGTTAAAAAGGGATCTGTCCCGTTTTCCTTCCGACCCACAGGGACACTG
FI:HOOH23:15 GGCGATGTGTACATCAATGATCTTCCTAGCCCTATGTCGTTGGTGAGGGTTGAGACTACATCTATACAGAGTAATACCTCCATGTATCCGCTTGAGCCAGTTCGGGTGGGAGTAGCGCCAGCACAACTGG
BRNV 23:2 GGCGATGTGTACATCAATGATCTTCCTAGCCCTATGTCGTTGGTGAGGGTTGAGACTACATCTATACAGAGTAATACCTCCATGTATCCGCTTGAGCCAGTTCGGGTGGGAGTAGCGCCAGCACAACTGG
FI:HOOH24:15 GGCGATGTGTACATCAATGATCTTCCTAGCCCTATGTCGTTGGTGAGGGTTGAGACTACATCTATACAGAGTAATACCTCCATGTATCCGCTTGAGCCAGTTCGGGTGGGAGTAGCGCCAGCACAACTGG
BRNV 24:1 GGCGATGTGTACATCAATGATCTTCCTAGCCCTATGTCGTTGGTGAGGGTTGAGACTACATCTATACAGAGTAATACCTCCATGTATCCGCTTGAGCCAGTTCGGGTGGGAGTAGCGCCAGCACAACTGG
510
FI:HOOH23:15 ACTATTGCCTATCTGACAATATTAGTTTTGCAAAGAGGGGTGCGAGCAGAAACACAACCCAAGTGAATAGGAAACGAATTGAAAGATATGGGAAATTGGAGGAAGTTGTCGGAGAAGATGAAGGGGGAA
BRNV 23:2 ACTATTGCCTATCTGACAATATTAGTTTTGCAAAGAGGGGTGCGAGCAGAAACACAACCCAAGTGAATAGGAAACGAATTGAAAGATATGGGAAATTGGAGGAAGTTGTCGGAGAAGATGAAGGGGGAA
FI:HOOH24:15 ACTATTGCCTATCTGACAATATTAGTTTTGCAAAGAGGGGTGCGAGCAGAAACACAACCCAAGTGAATAGGAAACGAATTGAAAGATATGGGAAATTGGAGGAAGTTGTCGGAGAAGATAAAGGGGGAA
BRNV 24:1 ACTATTGCCTATCTGACAATATTAGTTTTGCAAAGAGGGGTGCGAGCAGAAACACAACCCAAGTGAATAGGAAACGAATTGAAAGATATGGGAAATTGGAGGAAGTTGTCGGAGAAGATGAAGGGGGAA
FI:HOOH23:15 CGAGAGGAGAGGAAATTCCCACTCAGGAGATTCTTTCTGCGAGTCCCCACCTTGGGTCCACTATCCCGGTAGCCTACAACACTAATATTAAGGGCAAGGGTAAAGGCATTGAACCAATGGAACGAGTTAG
BRNV 23:2 CGAGAGGAGAGGAAATTCCCACTCAGGAGATTCTTTCTGCGAGTCCCCACCTTGGGTCCACTATCCCGGTAGCCTACAACACTAATATTAAGGGCAAGGGTAAAGGCATTGAACCAATGGAACGAGTTAG
FI:HOOH24:15 CGAGAGGAGAGGAAATTCCCACTCAGGAGATTCTTTCTGCGAGTCCCCACCTTGGGTCCACTATCCCGGTAGCCTACAACACTAATATTAAGGGCAAGGGTAAAGGCATTGAACCAATGGAACGAGTTAG
BRNV 24:1 CGAGAGGAGAGGAAATTCCCACTCAGGAGATTCTTTCTGCGAGTCCCCACCTTGGGTCCACTATCCCGGTAGCCTACAACACTAATATTAAGGGCAAGGGTAAAGGCATTGAACCAATGGAACGAGTTAG
667
FI:HOOH23:15 GATTGGCTCATCCTGGATTCGTGAGAGAAATGGGAGGAGGCGGCCTGAACCGATGTACCCGTATGGAAAACTTGAGAAATTGGATAATGCCCTGAGGAACGAACCGCAAACGGAGAAACACCATCATTT
BRNV 23:2 GATTGGCTCATCCTGGATTCGTGAGAGAAATGGGAGGAGGCGGCCTGAACCGATGTACCCGTATGGAAAACTTGAGAAATTGGATAATGCCCTGAGGAACGAACCGCAAACGGAGAAACACCATCATTT
FI:HOOH24:15 GATTGGCTCATCCTGGACTCGTGAGAGAAATGGGAGGAGGCGGCCTGAACCGATGTACCCGTATGGAAAACTTGAGAAATTGGATAATGCCCTGAGGAACGAACCGCAAACGGAGAAACACCATCATTT
BRNV 24:1 GATTGGCTCATCCTGGACTCGTGAGAGAAATGGGAGGAGGCGGCCTGAACCGATGTACCCGTATGGAAAACTTGAGAAATTGGATAATGCCCTGAGGAACGAACCGCAAACGGAGAAACACCATCATTT
FI:HOOH23:15 CAAGTGGCGGGGAATTGAGGCATCAGAGGGTACAGATTACCATGTGTGTTACAGGATTGGGCGGGCCCCATTCTACGGCTGTTGTGAACGCGATGAGATGCCTGTCAATATCATACCACTATCCATAAGA
BRNV 23:2 CAAGTGGCGGGGAATTGAGGCATCAGAGGGTACAGATTACCATGTGTGTTACAGGATTGGGCGGGCCCCATTCTACGGCTGTTGTGAACGCGATGAGATGCCTGTCAATATCATACCACTATCCATAAGA
FI:HOOH24:15 CAAGTGGCGGGGAATTGAGGCATCAGAGGGTACAGATTACCATGTGTGTTACAGGATTGGGCGGGCCCCATTCTACGGCTGTTGTGAACGCGATGAGATGCCTGTCAATATCATACCACTATCCATAAGA
BRNV 24:1 CAAGTGGCGGGGAATTGAGGCATCAGAGGGTACAGATTACCATGTGTGTTACAGGATTGGGCGGGCCCCATTCTACGGCTGTTGTGAACGCGATGAGATGCCTGTCAATATCATACCACTATCCATAAGA
FI:HOOH23:15 GAGGAAGCGGAGGAAAGCGTTAGACGGAGGGTCTTGGAGGCTCTCAATCCGGATGAGCTGTTTGAGCAATACTCCCACATGTATTCACAGGATGTTTTGATGGAAGCTTGCAGTCTAACCATACAAGATT
BRNV 23:2 GAGGAAGCGGAGGAAAGCGTTAGACGGAGGGTCTTGGAGGCTCTCAATCCGGATGAGCTGTTTGAGCAATACTCCCACATGTATTCACAGGATGTTTTGATGGAAGCTTGCAGTCTAACCATACAAGATT
FI:HOOH24:15 GAGGAAGCGGAGGAAAGCGTTAGACGGAGGGTCTTGGAGGCTCTCAATCCGGATGAGCTGTTTGAGCAATACTCCCACATGTATTCACAGGATGTTTTGATGGAAGCTTGCAGTCTAACCATACAAGATT
BRNV 24:1 GAGGAAGCGGAGGAAAGCGTTAGACGGAGGGTCTTGGAGGCTCTCAATCCGGATGAGCTGTTTGAGCAATACTCCCACATGTATTCACAGGATGTTTTGATGGAAGCTTGCAGTCTAACCATACAAGATT
1083
FI:HOOH23:15 ACCCTCAGGGGAACATTAAGTATGTGGACATGCTGGATGGAGTGCTCAAGGAATTCGGAGTGAACATCCGGGATTACAAGTACAGGGTGAACTGCGGTTGCTGTGGAAAGAAAGAGATGGTCCTCCAAT
BRNV 23:2 ACCCTCAGGGGAACATTAAGTATGTGGACATGCTGGATGGAGTGCTCAAGGAATTCGGAGTGAACATCCGGGATTACAAGTACAGGGTGAACTGCGGTTGCTGTGGAAAGAAAGAGATGGTCCTCCAAT
FI:HOOH24:15 ACCCTCAGGGGAACATTAAGTATGTGGACATGCTGGATGGAGTGCTCAAGGAATTCGGAGTGAACATCCGGGATTACAAGTACAGGGTGAACTGCGGTTGCTGTGGAAAGAAAGAGATGGTCCTCCAAT
BRNV 24:1 ACCCTCAGGGGAACATTAAGTATGTGGACATGCTGGATGGAGTGATCAAGGAATTCGGAGTGAACATCCGGGATTACAAGTACAGGGTGAACTGCGGTTGCTGTGGAAAGAAAGAGATGGTCCTCCAAT
1286
FI:HOOH23:15 GCAATAAAGGATGGCGCCTTAATGCACTTAAACGTGGTGGTTTCTCCGTGCGGGAAGGAAGAAGGACAGGCTTAGATTACCTCGGGTCACCATCATGGGCTGGCGGAGTTATTGGTATTACTAGCGATCT
BRNV 23:2 GCAATAAAGGATGGCGCCTTAATGCACTTAAACGTGGTGGTTTCTCCGTGCGGGAAGGAAGAAGGACAGGCTTAGATTACCTCGGGTCACCATCATGGGCTGGCGGAGTTATTGGTATTACTAGCGATCT
FI:HOOH24:15 GCAATAAAGGATGGCGCCTTAATGCACTTAAACGTGGTGGTTTCTCCGTGCGGGAAGGAAGAAGGACAGGCTTAGATTACCTCGGGTCACCATCATGGGCTGGCGGAGTTATTGGTATCACTAGCGATCT
BRNV 24:1 GCAATAAAGGATGGCGCCTTAATGCACTTAAACGTGGTGGTTTCTCCGTGCGGGAAGGAAGAAGGACAGGCTTAGATTACCTCGGGTCACCATCATGGGCTGGCGGAGTTATTGGTATCACTAGCGATCT
1414
FI:HOOH23:15 GACCCCAGTCGTCACACGGCAGGCCTTCGGAGAGCTAGTTGATCAGGTTGCCGAGAGCATGTTTGCTGAGAGCAAAAATAATGTACCTGGGTATCTCAAGGAGCTTGAGAAGTTCCACCC
BRNV 23:2 GACCCCAGTCGTCACACGGCAGGCCTTCGGAGAGCTAGTTGATCAGGTTGCCGAGAGCATGTTTGCTGAGAGCAAAAATAATGTACCTGGGTATCTCAAGGAGCTTGAGAAGTTCCCCCC
FI:HOOH24:15 GACCCCAGTCGTCACACGGCAGGCCTTCGGAGAGCTAGTTGATCAGGTTGCCGAGAGCATGTTTGCTGAGAGCAAAAATAATGTACCTGGGTATCTCAAGGAGCTTGAGAAGTTCCACCC
BRNV 24:1 GACCCCAGTCGTCACACGGCAGGCCTTCGGAGAGCTAGTTGATCAGGTTGCCGAGAGCATGTTTGCTGAGAGCAAAAATAATGTACCTGGGTATCTCAAGGAGCTTGAGAAGTTCCACCC